Contents

Executive Summary

Identifying risk, and doing so early, is a key component of the CBSA's mandate. One key method of risk assessment is through the Agency's Commercial and Traveller Targeting Programs. Targeting is currently supported by automated, *expert-based systems* in both the commercial and traveller streams.

Data mining in particular, uses machine-learning algorithms, with limited human intervention, to find faint patterns or relationships between data elements, in large, complex data sets, that are of business interest. At a high level, the use of data mining techniques

The Enforcement and Intelligence Directorate of the Programs Branch has worked closely with the Science and Engineering Directorate (S&E Lab) in ISTB to pilot predictive analytics at the CBSA, for pre-arrival risk assessment within the air traveller stream. The full scope of work for this pilot was completed within 10 months; 7 months were dedicated to preparing the data for predictive analytics, and 3 months were dedicated to advanced modeling.

Several machine-learning models were created and tested, with the best one giving very promising results. By making use of a relatively small amount (i.e., only API and PNR data) of the information available to targeters in their decision-making process, it was possible to correctly predict up to of the *known* illicit cases. These results were achieved

The results achieved can thus be seen as a lower bar as to what is possible.

The S&E Lab will proceed to phase two of this pilot in an attempt to address some of the more difficult aspects of this work and to improve the already promising results. Although continuing this work will require a concerted effort from many different parties, the promising results of this pilot suggest that this work will be well worth the effort. This pilot has proved the value of this paradigm shift, and that devoting a focused effort to this type of work could lead to enormous value for the Agency. Moreover, these techniques are applicable to many other of the CBSA's lines of business

and so these initial efforts can also be seen as a powerful opportunity to educate the Agency at large, as to the power and value of such work.

Background

Identifying risk and doing so early is a key component of the CBSA's mandate. One key method of risk assessment is through the Agency's Commercial and Traveller Targeting Programs. These programs rely on

Targeting is currently supported by automated, *expert-based systems* in both the commercial and traveller streams.

by using predictive analytics techniques, the Agency can sift through large data in an automated fashion,

Predictive analytics includes advanced mathematical and statistical techniques, which look at historical data to make predictions about future events. The goal of such work is to use machine-learning algorithms, with limited human intervention, to find faint patterns or relationships between data elements, in large, complex data sets, that are of business interest. Because the analysis does not start with a specific hypothesis , the algorithms may unbiasedly identify trends that would otherwise go undetected. At a high level, the use of data mining techniques

Given the large volume of travellers and conveyances coming to Canada, along with the ever increasing collection of data, and rapidly developing technologies, the Canada Border Services Agency (CBSA) must begin to leverage predictive analytics to assist in the automation of risk assessment.

Exploratory work in the field of predictive analytics was carried out by the Science and Engineering Directorate, in the Information, Science and Technology Branch (ISTB), between 2010 and 2013, in the

The Enforcement and Intelligence Directorate of the Programs Branch has worked closely with the Science and Engineering Directorate (S&E Lab) in ISTB to pilot predictive analytics at the CBSA, for pre-arrival risk assessment in the air traveller stream. The full scope of work for this pilot was completed

within 10 months; 7 months were dedicated to preparing the data for predictive analytics, and 3 months were dedicated to advanced modeling. This report describes the methodology used by the AA section, the results obtained through machine learning techniques, and the applicability of machine learning at the CBSA.

Methodology

Data Extract: Passenger Information System (PAXIS) and Integrated Customs Enforcement System (ICES)

In order to facilitate risk assessment in the traveller stream,

(the tools used to access and analyze the data).

In order to pilot a predictive analytics in the air traveller stream, the S&E Lab required access to the **same** data described in the previous paragraph. The team

However, due to the legal and technical requirements of the *Protection of Passenger Information Regulations* (PPIR) and the commitments made to the EU, the data extract request was modified at the request of the data owners (Programs Branch) to exclude the transfer of As a result, the Enterprise Data Warehouse Division in ISTB extracted from the production environment onto the Enterprise Data Warehouse (EDW), the week of June 15, 2015, and access to this data was given to the S&E Lab.

Data Matching: Matching Records from ICES to PAXIS

The development of a predictive model depends heavily on matching the this is because data mining requires information about a passage

. This means that a small set of passages will be classified as resultant (i.e., all entries in ICES that correspond to an entry in PAXIS). **All other records** in PAXIS that do not contain a corresponding record in ICES will be categorized as non-resultant.

In an effort to link the ICES data to its corresponding API/PNR data the following set of rules were applied by the EDW Division. The rules were applied to data from three tables in PAXIS and one table in ICES.

If all eight rules above, matched an ICES record to a PAXIS record, then this record would have qualified as a match; otherwise no match would have been made. Note that of the seizures recorded in ICES for the period of December 1, 2013 to May 31, 2015, matches were made.

Given that by this point, much work had already been done, the S&E Lab did not ask the EDW section to implement any relaxed criteria. A maximum of extra matches would not have provided enough of an improvement compared to the amount of work that would have been required to implement such changes.

The figure below shows, very simply, that matching rules were created and applied to link up records in ICES with their corresponding API/PNR data.

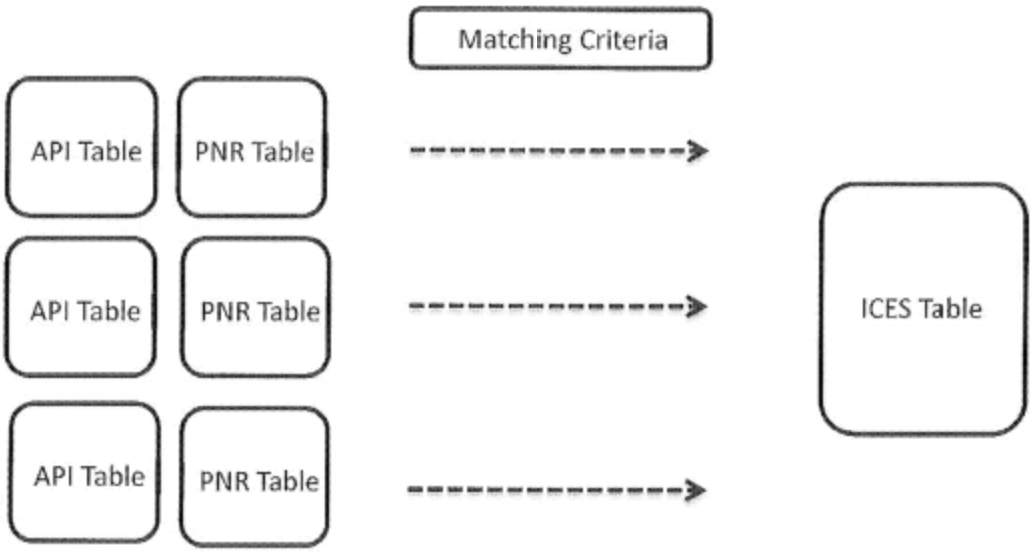


Figure 1. Matching process between API, PNR and ICES datasets.

Data Reduction: Reducing Dataset to Unique Passages

Final Data Preparation Step

The previous step was required in order to identify all of the data elements from the PAXIS side that corresponded to a particular seizure on the ICES side. All travellers recorded in ICES and matched to a record in PAXIS based on the rules described in the previous section, have several corresponding primary keys These primary keys, in theory, should allow for the ICES and API/PNR tables to be stitched together, to end up with a large dataset such as the one that is pictorially oversimplified below.

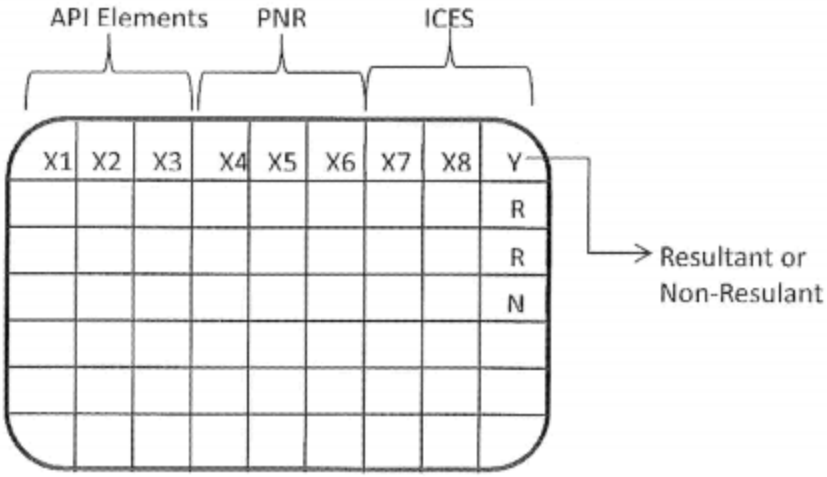


Figure 2. The different datasets linked together into a complete, full dataset.

However, the previous process is complicated by the following relationships:

As a result, logic had to be formulated in order to reformat the data, so as to end up with one row for each unique passage into Canada. Two IBM Modeler streams were independently created so as to confirm the logic of one stream with that of the other, and vice versa. This was done as a way to validate the logic prior to model building, and as a way to build in error and quality checks. Both streams follow very similar (almost identical) logic and rules. Once the data preparation was completed, the model-ready dataset contained:

- Approximately 37 million unique passages into Canada (which falls in the range that is reported by the CBSA, in the Consolidated Management Reporting System)
- Approximately (data from the API/PNR and ICES tables)
- Roughly of the above 37 million passages corresponded to travellers that were caught with contraband

The data preparation steps described above were very challenging and time consuming activities (~ four of the six months allotted for the pilot);
Future iterations of this pilot would require the above data preparation logic to be repeated, but would be more streamlined.

Graphical Exploration of the Model-Ready Dataset

At this point, some descriptive statistics will give some context to the raw data that was used to create predictive models. Below, some basic statistics are reported for the full ICES results from the 18-month period of the data for this pilot, as well as for the subset that had matched to API/PNR data. Table 1, below, shows the top ten types of resultants within the dataset;

Table 1. Top 10 seizures within the model-ready dataset.

Type of Resultant	Percentage	Count
Currency/Monetary Instrument		
Tobacco Products		
Clothing & Footwear		
Jewellery		
Other Controlled Drugs		
Furs, Skins, & Leather Products		
Marihuana		
Alcoholic Beverages		
Watches		
Cocaine		
Subtotal		
Others Resultants		
Total	*100.00%*	

Figure 4 below, highlights that the contraband caught, follows

This figure also shows that the number of seizures made per month, follow a distribution.

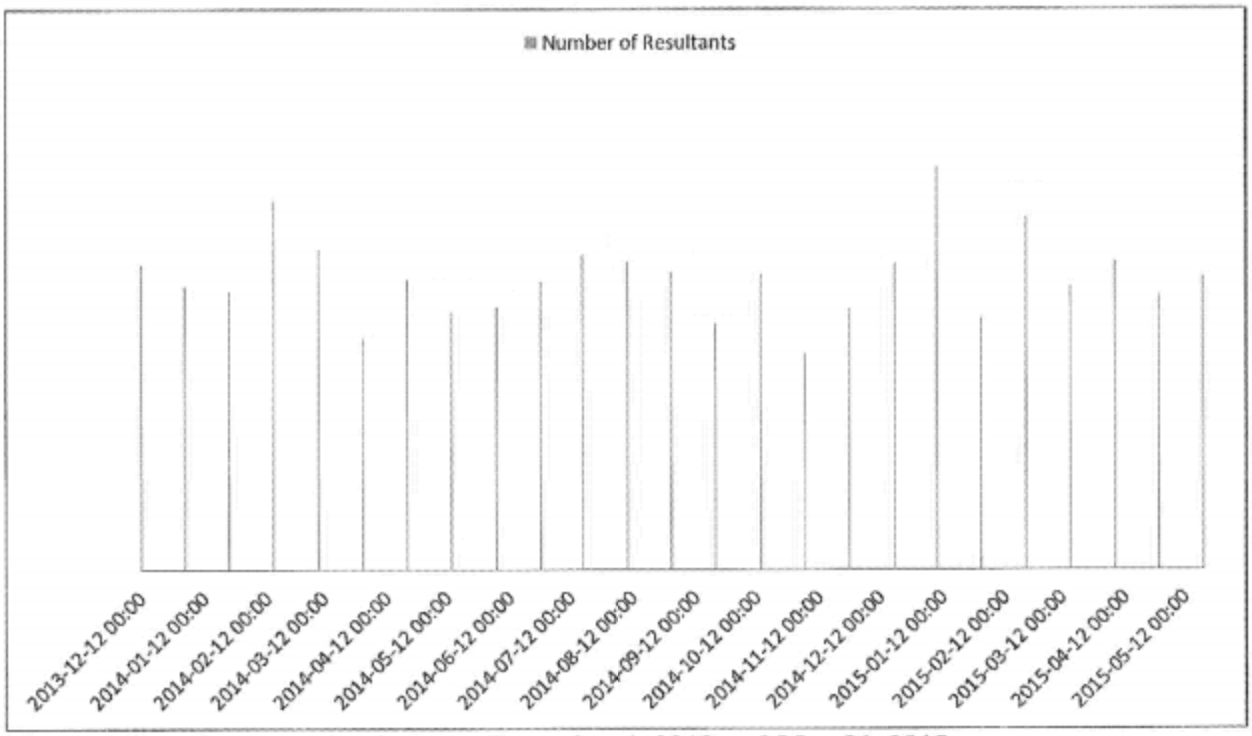


Figure 3. Resultant matches between December 1, 2013 and May 31, 2015.

Table 2 below shows that (in the model-ready dataset) were reported (in ICES) as being referred by a BSO. This trend is consistent when all seizures are considered (for the 18-month period), as seen in Table 3. Digging even deeper still, of all resultants in the same period were associated with in ICES.

Table 2. Contraband seizures in the model-ready dataset, by referral type.

Type of Referral	Percentage	Count
Total	*100.00%*	

Table 3. Total seizures between December 1, 2013 and May 31, 2015, by referral type.

Type of Referral	Proportion	Percentage	Count
	1	100.00%	

Variable Creation

Next, the model-ready dataset described above, with the raw data variables (i.e., variables in their raw form pulled directly from API/PNR, and ICES tables), was used to create new elements for use in the model building phase.

Some variables in the dataset, in their raw state, are unusable for modeling. A prime example of such a variable would be any On their own, are not usable for modeling purposes, because they don't say much in and of themselves. However, in comparison with other in the data, they could be quite valuable.

The S&E Lab focused their efforts on the creation of data elements that could lead to interesting patterns for model creation. An opportunity exists for different areas within the CBSA, to get involved here, particularly those responsible for creating the scenarios for Scenario Based Targeting (SBT). With their extensive knowledge and experience, they may be able to suggest potentially useful information and thus help to create potentially useful variables as inputs to the modeling step, in future iterations.

Modeling – Training and Testing

At this point, the dataset was ready to be modelled. The first step in the model building phase requires splitting the dataset up into two sets: one for "training" and one for "testing". A training set is a subset of the full data, which is used to create a model. A testing set, is the other subset of the data (what is left over), which is used to test the model that was created in the training set. The reason the data is split in this fashion, is because it is important to test the model using new data (i.e., data that the model has not "seen", or which has not been used to create the model).

The data was split so that one year of data was used for training, and the last six months was used for testing. Splitting the data in this manner allows for the results to be used as a true simulation of what would have happened had the models created been implemented in production on the 366th day, and run for the following six months. Of the passages, approximately were used to train models, and were used to test the models that were built. In the training set, of the records, there were roughly resultants, and in the testing set, of the records, there were roughly resultants.

Several different machine-learning techniques were used to build models with the training dataset; these include decision trees, decision lists, logistic regression and neural networks. These models were all generated without human intervention and their outputs are solely data driven; the data drives the patterns that are found by the algorithm. Due to the time constraints in this pilot, most of the efforts were spent on outputting a model that would be interpretable and that would showcase the potential for straightforward integration into current targeting systems (SBT). As a result, many different decision trees were built. One in particular gave the best results, in comparison to all models built.

A **decision** or **classification tree** is a predictive model that acts as a decision tool and is straightforward to understand and follow. In short, a decision tree uses the observations about a particular item, and relates it to an outcome. In this pilot for example, a decision tree maps several observations about a particular passage (i.e., API/PNR data from a particular passage) and relates it to a risk category (i.e., low risk/high risk, or non-resultant/resultant). Along with the prediction of risk level, a decision tree outputs a confidence measure in the predictions; these confidence values can subsequently be used to determine thresholds that would highlight which passages and passengers to focus on.

Depicted below is an excerpt from the best model that was generated. The full tree was too large to be presented here.

Figure 4. An excerpt of the best model (a decision tree) built and tested.

An example of how this part of the tree can be read, is as follows:

-

Results

Best Model – A Decision Tree

This section describes how the best model was assessed, as well as the implications for the Agency. Recall that the testing set is a subset of data that is used to assess how well the model that was created, performs. The test set that was used, included approximately passages into Canada, of which were resultant (i.e., passages that corresponded to an ICES seizure for contraband). The general results of this work are depicted graphically below.

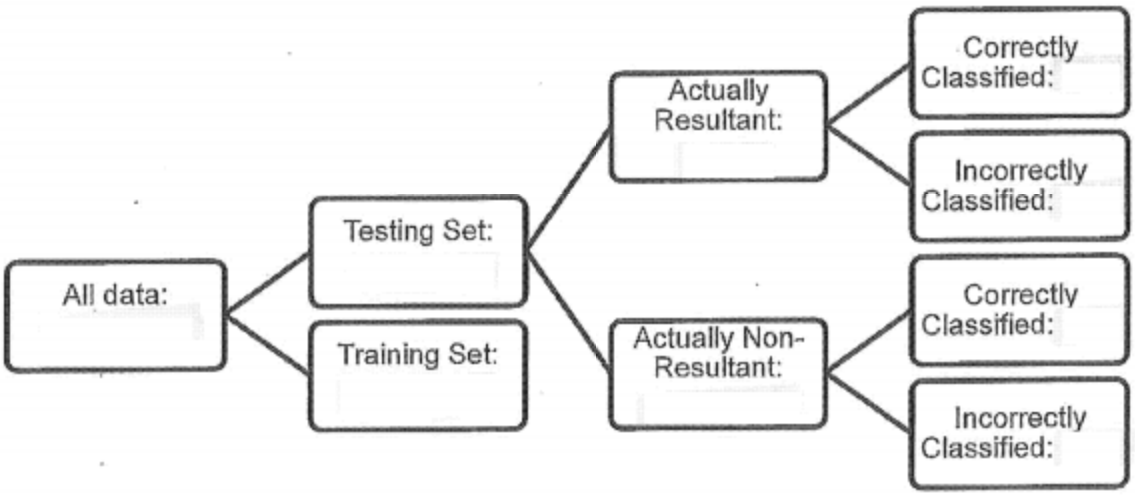


Figure 5. Assessment of the best model that was created.

Put into matrix form, the above results can be summarized as follows:

Table 4. Confusion matrix showcasing the results of the best model created.

		Predictions	
		Predicted Non-Resultant	Predicted Resultant
Actuals	Actual Non-Resultant		
	Actual Resultant		

The results above show that the best model that was built, can correctly predict of the non-resultants and of the resultants. That is,

This has significant implications for the Agency.

A perfect model would predict 100% of the non-resultants as non-resultants, and 100% of the resultants as resultants. So, in the above case, the off-diagonals (in red) are the errors; of the resultants were incorrectly predicted as non-resultants and of the non-resultants were incorrectly predicted as resultants. If this model were implemented solely on its own, without additional intelligence or BSO intervention, For example, implementation of this model, alone, would result in roughly The obvious question that follows is, "How big is More importantly, "*Would pose a problem operationally?*" If it *is* considered too large of a number, there are ways to reduce it (which will be discussed next). However, relying only on the decisions of a model would not be advisable at this stage.

It is also important to note that these results do not imply that all incorrectly classified cases of contraband smuggling would have gotten away had this model been used in the CBSA's production environment. One has to remember that these cases were caught using the combined strength of the CBSA's interdiction efforts That said, what the above results are truly showing is that, using this initial model, in only of cases was the risk of contraband smuggling prior to the traveller's arrival in Canada not identified.

Model Based on a Threshold of Confidence

Given the results in the previous section, the goal of this next step was to create a more feasible and realistic model solution

To this end, the S&E Lab decided to employ a fairly simple technique: every 'resultant' prediction with a corresponding confidence value equal to would remain a 'resultant' prediction; all other 'resultant' predictions would at this point be re-categorized into a 'non-resultant' prediction.

The table below shows that this process reduced the false positive errors (from down to , however, in doing so, the false negative errors increased (from The positive point to highlight is that the thresholded model, if implemented on its own, would now result in roughly instead of More importantly however, is the ability of this model to still pick out of the resultants, using only two sources of data, and doing so pre-arrival.

Table 5. Confusion matrix showcaseing the results of the best model, thresholded (based on high confidence predictions).

		Predictions	
		Predicted Non-Resultant	Predicted Resultant
Actuals	Actual Non-Resultant		
	Actual Resultant		

Assessing Results in a Business Context

The metrics used in the previous section to assess the performance of the statistical model are informative, but do not easily highlight how the model could impact the CBSA's business processes. As mentioned, The problem with this line of thinking is two-fold:

1. Thresholding requires weighing the cost and benefits of relying only on the best possible "risky" predictions (i.e., there could be many accurate predictions occurring below the selected threshold).
2. The intention of this pilot project is not to say that that there would be That said, a larger amount of predicted resultants might be acceptable, given that the predictions would only serve to populate a worklist.

With these points in mind, it could be argued that the resulting from the predictive model could actually be more operationally feasible than thought to be, at first glance. To belabor the point,

and was thus impossible for the purposes of this pilot.

The following diagram attempts to highlight the difficulties in comparing the work presented in this report, against the CBSA's current targeting process. The figure below points out that the common evaluation metrics,

the figure below should highlight the fact that these two levels of evaluation are quite different.

Figure 6. Comparison of current targeting process against the predictive analytics pilot process.

The most appropriate comparison between the CBSA's current process and this pilot would look at the

Due to this fact and in attempt to ensure that proper comparisons were made to fairly evaluate this pilot project, only self-contained statistical measures were relied upon. The following section will present another such metric that will point to the efficacy of the methods described.

Assessing Results against a Random Process

A standard evaluation metric in almost any data mining project evaluates the performance of the created models against the performance of a random selection process. Clearly, this method of evaluation is less informative than comparing a model against the statistics generated by the current targeting process. However,
comparison to random *does* provide an effective and self-contained way of assessing the performance of the models.

The current targeting process and the process generated in this pilot, serve as different levels of screening, which in turn filter through the data received to identify records of interest. Without any levels of screening, the only option left to identify records of interest would be a random selection process; this can be considered a baseline from which to make a fair comparison. Clearly, if any screening process, be it traditional targeting or statistical modelling, performed worse than random, there would be no use in relying on the decision-making ability of the system. That said, in using a random process as an evaluation metric, the ideal result is to find improvement, or "lift" from the random baseline.

In this pilot, a random baseline can easily be established. Recall, the six-month testing set was composed of approximately records, close to of which were resultant for contraband. If a record were picked at random from the test population, there would be a likelihood of that record being a resultant case. See below for the exact calculation.

1. **Choosing at random within the test set:**

$$\frac{Resultants\ in\ Test}{Passages\ in\ Test} =$$

2. **Using the model:**

$$\frac{Correctly\ predicted\ resultants}{Total\ predicted\ resultants} =$$

better than random)

3. **Using the model with high confidence resultants:**

$$\frac{Correctly\ predicted\ resultants}{Total\ predicted\ resultants} =$$

better than random)

Similarly, once the test data is run through the statistical model, records predicted resultant, This means that of all records predicted resultant by the statistical model, there is a of a given record being resultant. This is better than

Finally, if the test data is run through the thresholded statistical model, out of the predicted resultant, were truly resultant. This means that of all records predicted resultant by the thresholded model, there is a of a given record being resultant. This is better than the performance described above.

but it does provide a baseline. With the models performing better than it is clear that there is potential benefit to employing these sorts of data mining techniques. Moreover, when paired with the fact that many details have been identfied that could greatly improve this preliminary work, the comparison made in this section could be seen as a strong reason to continue this analysis, as discussed in the next section.

Conclusions and Future Work

The goal of this pilot was to investigate the efficacy of using machine-learning techniques on pre-arrival data in order to identify cases of probable contraband smuggling.

it was indeed possible to correctly predict up to of the known illicit cases.

Moreover, it is worth restating that these results were achieved in the face of certain obstacles that most certainly made this work much more difficult.

The S&E Lab proposes that this work proceed with a follow-up round of analysis in attempt to address some of the more difficult aspects of this work, and to attempt to improve the already promising results. An additional benefit to another round of analysis is that it will provide the opportunity to make a more concerted effort in terms of gathering statistics

Perhaps the main point of improvement for the next phase of this work has to be centred on the

For example, are collected by the CBSA at the border through IPIL. Each traveller entering Canada has his/her passport scanned, generating data regarding the individual in question, which is then saved in CBSA databases.

This information

In order to advance this work with a second round of analysis that includes the improvements previously mentioned, the S&E Lab will require continued assistance from various areas of the Agency. Assistance will be needed from (but not limited to):

1. **Operations Branch:** Work with NTC subject matter experts for variable creation and perhaps API/PNR data preparation.
2. **Programs Branch:** Support required for the Round 2 data extract, including extract of new fields.
3. **ISTB:** Support from EDW for the Round 2 matching process, and possibly flattening of the data.
4. **Data Analytics Work Group:** Support and direction in terms of socializing the concepts and value of this type of work.

The work described in this report can truly be seen as a lower bar as to what is possible. That said, although continuing this work will require a concerted effort from many different parties, the promising results of this pilot suggest that this work will be well worth the effort. This study has proved the value of this paradigm shift, and that devoting a focused effort to this type of work could lead to enormous value for the Agency. Moreover, these techniques are applicable to many other of the CBSA's lines of business and so these initial efforts can also be seen as a powerful opportunity to educate the Agency at large, as to power and value of such work.

Canada Border Services Agency | Agence des services frontaliers du Canada

Protected B

Science and Engineering Directorate

Border Technology Division

Division Report 2015-01
February, 2015

Deliverable #1 – Use of IBM SPSS Modeler for Text Mining of Cargo Descriptions.

Darren Coughtrey

PROTECTION • SERVICE • INTEGRITY

Canada

Executive Summary

Currently, prior to the arrival of goods to the country, the only description of the physical goods that is received is provided in a free-form text field. This study uses commercial off the shelf products that the Agency has made a major investment in, to be able to extract standard commodity descriptions from the free-form text in order to be able to use it in an automated fashion. The results that are presented are compared with previous work performed by the Advanced Analytics Section, and suggest that actionable information can indeed be extracted from this textual information

Purpose

In order to both secure the country's borders and facilitate legitimate trade, many attempts are being made to automate historically manual processes overseen by the Agency. One of the most critical processes being automated by the CBSA is the pre-arrival risk assessment of commercial cargo bound for Canada. The Agency has been working towards automation of the risk assessment process since it's inception in 2001, which has resulted in the creation of TITAN, an automated system aimed at aiding the human targeter's search through the waybills of incoming freight conveyances in both the Marine and Air modes.

TITAN (to be renamed 'PHOENIX' by the completion of e-Manifest) has been continually improved upon since its initial implementation in 2004, and upon its completion, the eManisfest project will result in TITAN being extended to both Highway and rail Modes. Moreover,

[1] Division Report 2012-11 (TR)

An important point to note within the context of this report is that many risk rules, such as the one described in the preceding paragraph, are relatively trivial to implement.

the creation/implementation of these rules really only requires the Agency experts to maintain a list of the ISO Codes for being flagged by the rule in question. Because this information is so easy to input, the tends to be transmitted very consistently and error-free by the trade community. Paired with the fact that comparing incoming ISO Codes with those being flagged by the CBSA experts is an operation that is trivial to implement, it is safe to say that these sorts of data elements form a solid basis for risk rule creation.

However, with such a variety of risks under the Agency's purview, not all of the risk rules created for use within the automated risk assessment process run on clean, codified data.

In comparison to the more straight forward rules based on rules based on text present many problems in terms of automation;

[2] http://www.iso.org/iso,

Many of these questions have been well documented and analyzed[3]. With the exception of a previous report[4] generated by the Science & Engineering Directorate's Advanced Analytics Section, much of the work done in this area has centered on identifying the issues, as opposed to posing solutions to them. The following sections, however, will outline a potential solution using the technology readily available within the CBSA. Firstly the methodology will be outlined at a high level, followed by an analysis, and presentation of the results.

An important aspect of this study was to make use of the text analytics functionality within the IBM SPSS Modeler data mining suite. In recent years, the Agency has made a significant investment in this technology, and made an additional purchase of a component specifically aimed at making use of textual information in an automated fashion. In the future, the CBSA may be able to make use of IBM suite, which would allow work done in Modeler to be published and used directly in the Agency's production systems. Viewed in this context, the results described herein may represent the horizon of the Agency's risk assessment capabilities.

With regards to textual data, the aforementioned software is aimed at allowing for the creation of statistical models that could highlight patterns and trends within the text, which could then be used for automated tasks like classification (i.e., categorizing the text into different categories), or sentiment analysis (i.e., attempting to extract a sense of what the text is saying about a given topic). In order to help with these tasks, the software provides various built-in subject specific dictionaries, and text mining algorithms that can be used in conjunction with the data manipulation operations and data mining techniques that are built into the more standard portions of the suite.

Methodology

To begin with, one must understand the high-level processes that get used with the IBM SPSS Modeler's text mining functionality. The first thing to understand is the software's use of dictionaries; these are predefined inputs to the text mining process, which essentially define the words recognized by the text mining software (i.e., any word not appearing in the chosen dictionary will be ignored in the subsequent processing). That being said, the term 'dictionary' seems a little less appropriate than using the term 'vocabulary' for the chosen lexicon. Nevertheless, the software comes with several preinstalled dictionaries, many of which are subject specific (banking, marketing, etc.), and not

[3] Division Report 2013 - 01(TR), Division Report 2014 - 02(TR)
[4] Division Report 2013 - 03(TR)

necessarily applicable for the CBSA's purposes. For this reason, this study makes use of the standard English dictionary provided with the software.

The software does give the user the ability to add any domain specific language to the lexicon that may be lacking. However, for reasons that will be described in the discussion after the presentation of the results, doing so for this particular piece of work would have most likely resulted in a waste of effort in that the number of terms needed to be extracted already overwhelms the software as is.

Despite these potential set backs, what will be shown is that there is potential for optimism when dealing with these textual cargo descriptions. The results given will show that there is indeed potential to extract enough information from the text in order to give a more standard description of the commodity the text is attempting to describe. In pairing this with previous work[5] performed by the Advanced Analytics Section, there is reason to view this particular problem as solvable with the proper technological framework in place.

To begin, a certain amount of data pre-processing was required. In attempt to attain some form of standardization amongst the textual descriptions, two broad data cleansing steps were undertaken. Firstly, all numerical characters were stripped from the descriptions, simply because, if provided correctly, the commodity should be described in the text by words. Indeed, numbers may appear for the purposes of relaying the quantity of the goods being described, but the commodity itself should be present as text string somewhere in the description.

The next pre-processing step was to remove all punctuation characters from the text. This step might seem somewhat dubious, in that IBM SPSS Modeler does provide the capability for some level of Natural Language Processing, which would make use of punctuation, in order to tag the parts of speech within the descriptions (tagging the nouns, adjectives, names, etc.). Due to past experience in using a Natural Language Processing package on similar descriptions, the removal of punctuation was deemed a justifiable step. The problem that these descriptions are rarely properly formed sentences and are often merely just lists of nouns, making any effort in performing Natural Language Processing largely a waste of time and computation.

Once these basic data cleansing operations were performed, the data was then fed through the IBM Text Mining module. As mentioned above, the module makes use of the selected dictionary (i.e., the basic English dictionary in this case) to weed out words (from the text) that don't appear in the vocabulary, and to perform several more advanced processing steps (identifying synonyms, finding multi-word concepts, etc.).

An important point to note is that the Text Mining functionality within the IBM suite gives two ways of generating a text mining model for use within a data mining stream; the first is aimed at novice users and allows them to generate a model directly with little user-interaction, while the second method is interactive, and requires the user's input to generate the model. A key point in this study, with regards to these two options, is that the more simplistic option allows for the extraction of only 1000 concepts, and the more complex, interactive option extracts up to 10,000 concepts. In both cases, these

[5] Division Report 2013 - 03(TR)

'concepts' are typically single words, or perhaps two-word groups that occur commonly together, such as "car parts", or "apple juice", etc. In order to foreshadow problems to be discussed later on, consider the fact that there are one million possible standardized codes for commodities in the United Nations Harmonized System at the six-digit level alone.

That aside, once the concepts were extracted, the resulting model was then used to turn the data into a so called 'Bag of Concepts'. This means a new variable was created for each concept, which took on the value of 1 or 0 depending on whether or not, the given concept appeared in each piece of text. It is critical to note that this process dramatically increases the size of the data, and as will be shown, can be burdensome on memory requirements. Prior to creating the 'Bag of Concepts', the data set had a single text-based data element, whereas at this point in the process, the data set would potentially have 10,000 different binary variables.

That being said, once the 'Bag of Concepts' is generated, the analysis and data manipulation techniques included in the base module of IBM SPSS Modeler can be used to extract patterns and trends to use for predictive purposes. This means that standard data mining algorithms can be used to produce predictive models like decision trees and neural networks. A particularly useful type of analysis for high dimensional, binary data is what is known as 'Market Basket Analysis', which can be done within IBM data mining suite using the Apriori algorithm. This particular algorithm was discussed in a data mining workshop session[6] put on by the Advanced Analytics team, and so in what follows, familiarity with this concept, is assumed, but at a very high level. The relative rarity/commonality of the different concepts in the data (which is used to train the model) is used to create association rules that have a guaranteed predefined confidence and support level. That is,

If **A->B** is a rule generated by the algorithm:

1. The confidence of the rule will be the number of descriptions in which **A & B** appear divided by the number of descriptions in which **A** appears. This measures the strength of the rule.
2. The support of the rule is the number of times **A & B** appear together in a description divided by the number of descriptions. This measures the rarity/commonality of the rule.

Note that in many applications, rare rules are often seen as non-useful, and so the threshold for the minimum support of any rule is set relatively high in order to ensure that the conclusions drawn from the rules occur relatively frequently in the training data. In this context, however, rare rules are potentially useful. Keep in mind that the goal in this study is to take any description, regardless of wording, and extract a standard form for the commodity being described. That being said, if certain words are used, even infrequently, to describe a certain product, it could be beneficial in being able to account for the immense variability in the ways that any product can be described.

With this in mind, it is also useful to note the modeling software allows the user to restrict the conclusion of the rules generated by the Apriori algorithm (i.e., the right-hand side of the rule, or **B** in

[6] Data Mining Workshop Series - # 5

the case of A->B). That being said, if a standard way to describe commodities can be identified, the software can then be used to create rules of the following form

'Concept 1' & 'Concept 2'&...& 'Concept N' -> 'Standard Definition for Commodity Y'

That is, by using the concepts extracted from the textual descriptions, it may be possible to infer a standard definition for what the text is trying to describe. This would eliminate the need of any troublesome, fuzzy matching in the automated risk assessment processes; that is, the Agency could then compare standard text to standard text.

Fortunately for the purposes of this study, the UN's Harmonized System does give standardized definitions to any commodity, and on the Release documentation collected the CBSA collects both a free-formed text description, and a standardized definition in the form of a Harmonized System Code (HS Code). This means that this data can be used to generate a rule set of the form described above. The following sections will describe the results of this undertaking, and attempts to interpret them in the light of previous work. They will also highlight the difficulties of this process, and give reason to believe that they may not be insurmountable.

Results

Indeed, issues with this approach arose almost immediately. Recall that rare rules are potentially valuable in this study, which means that the support (and even confidence) thresholds can, and perhaps should, be set low in order to cast as wide a net as possible. Initially, however, this would result in massive rule sets that would be comprised of some 10,000 to 20,000 rules. In and of itself, a rule set this size is not an issue. In fact, given the myriad of possible ways to describe a single product, it might even be expected.

Indeed, this is exactly what happened in this experiment. The data was split into two groupings, one from 2012 and one from 2013, and then 200,000 lines were selected randomly for each set. This data was pre-processed as previously described above, and Apriori models, described above, were trained

on the data from 2012. The resulting rule set was used on the data from 2013 in order to test it's the ability to generate the correct HS Code from a free-form text field.

Initially the generated rules were run against the 2013 data, and only one prediction was generated per description. Recall, there is nothing stopping multiple rules from hitting on a single description. So, to generate a single prediction, one needs only to take the prediction given by the rule with the highest confidence. Subsequent predictions can be made by taking the results of the next highest rule hits. That being said, with only one prediction being generated,

In examining the resulting predictions, it was noted that a significant number of cases had no rules fire against them, and thus no prediction was generated for such cases. If such cases were excluded, the accuracy of those left over That is to say, for those cases for which the resulting model was able to generate a prediction, the model predicted correctly with only a single prediction.

Now recall, for no rule to have fired on a particular case, it would mean that either the parameters for the model building were set too high (i.e., rules involving the concepts present in the given description did not meet the minimum support/confidence criteria), or the concepts present in the given description were not extracted when the text mining model was built. While it is possible that concepts involved were so rare that they did not appear in the 2012 training data, the more likely explanation is that the

Now, to help improve the results, and in line with the previous Naïve Bayes work, other results were generated by outputting more than one prediction per case. As in the previous study, it can be argued that doing this is still sufficient for risk assessment purposes, because in effect the model would be saying 'out of the huge number of possible commodities, I have narrowed it down a small number of possibilities, and one of them is correct with X% confidence'. The following table summarizes the results of generating up to three predictions.

	Number of Predictions Generated		
	1	2	3
All Cases included			
Cases with no Prediction removed			

What this means is that with only two predictions being generated, it is possible to correctly predict

By adding one more prediction, these results are improved another
respectively. These results are on par with previous work, and most likely better for the case of generating a single prediction. This fact may be further proof of the efficacy of more complex concept extraction techniques, because although the previous work had no memory limitations, it perhaps suffers by having a very simplistic concept extraction process. Nevertheless, when taken together with previous results, this analysis does seem to confirm that it is possible to extract actionable information from this free-formed text that has caused such problems for the Agency's automated risk assessment processes.

Conclusion

Although it may not be obvious at first glance, this work can be viewed as quite positive. Moreover, this study lends credence to earlier work done by the Advanced Analytics Section. What was shown herein is that, although this initial attempt at using the IBM SPSS modeler software to process these troublesome text descriptions may not have been exact, it is possible to exploit the COTS products already purchased by the CBSA to dramatically prune the HS hierarchy into a small list of choices for a significant portion of incoming shipments.

The thing to remember here is that the limiting factor to this work and its results
not the techniques themselves. There is reason to think that results would drastically improve if a larger number of concepts could be extracted. In fact, in previous work, the Advanced Analytics Section was able to use the Naïve Bayes algorithm to great effect in trying to predict the six-digit HS Code from text[7].

Previous work also took the equivalent of one-word concepts in the IBM context, and so the work described in this report could even be seen as proving the benefit of being able to extract more complicated multi-word concepts, which was in fact a hypothesized improvement to the previous Naïve Bayes work. The point being, that if a bigger dictionary of multi-word concepts could be generated, it is possible that the correct commodity could be extracted from the text, accurately, with only the ideal of one prediction made.

The one advantage the earlier Naïve Bayes work had over the study being described herein was that it did not suffer from That being said, its allowable vocabulary was massive (i.e., a whole English dictionary with domain specific words added in), and the resulting model was able to accommodate millions of variations in the text.

[7] Division Report 2013 - 03(TR)

PROTECTED B

MEMORANDUM

To: **Phil Lightfoot, Sofia Auer**
From: **Darren Coughtrey**
Subject: **Using Textual Data in an automated way in the CBSA**
Reference: **DC02112015**

Introduction

Over the past several years, the Science and Engineering Directorate's Advanced Analytics team has amassed a large number of results directed at

These various projects have highlighted several overarching points that will be tied together in this memorandum. To begin with the reader must be aware of the problem that the AA team's work was aimed at addressing.

It is a well known fact that all commercial shipments bound for Canadian soil through Marine and Air travel

Many of these risk rules are straightforward because of the nature of the incoming data.

That being said the process of matching incoming data to the maintained list is a simplistic exact match on the codified elements.

There are, however, different rules that add new layers of complexity for the Agency's automated systems due the fact that the data format does not as easily handled

The prime example of this type of rule and the focus of much of the AA team's work is the data element giving the description of the commodity being shipped to Canadian soil. Unlike codified elements described above, this particular data element comes in the form of a free-form text field with essentially no standardization. The trade chain partners are basically free to enter into this field whatever information they see fit, and unfortunately for the CBSA, they do.

Things like contact information, internal company coding, and short forms that verge on gibberish appear in the data, and it falls to the CBSA to sort this out in order to be able to make a risk determination.

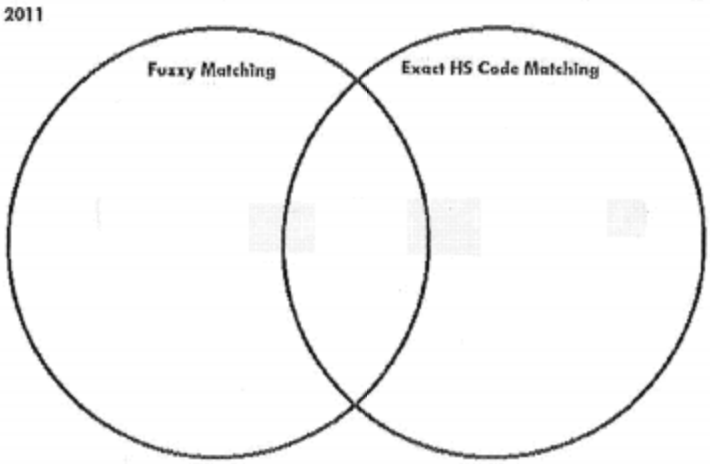
The unfortunate thing, however, is that fuzzy matching on text is an *extremely* difficult task. It can be done in fairly naïve ways that aim to account for simple spelling mistakes (e.g., omitted or swapped letters), or it can be done with fairly sophisticated techniques that take into account context, structure, and grammatical nature of the words in the text.

Moreover, the difficulties with fuzzy text matching are obviously exacerbated by the aforementioned data quality issues. In a nutshell the problem is that with the myriad ways one can legitimately describe any given commodity, and knowing that the text the Agency receives is of poor quality and often contains superfluous information; what words are the CBSA's risk assessment experts to put on their tables of risky commodities?

The following sections will briefly outline the various pieces of work conducted in this regard, and will point to the associated report for the interested reader.

Risk Rule Performance

This was done because the Release sent by the importers themselves includes the full ten-digit HS Code, and because the importers face steep penalties for non-declaration or misdeclaration, this data could serve as a baseline for the analysis of text-based rules. That is, under the seemingly reasonable assumption that the received HS Codes are by and large correct, one can translate the risk rule tables from text to HS Codes, and then calculate how many shipments *should* have caused a particular rule to fire.



Typically as well, it was found that the text-based rules fired far more often than risky commodities were declared on the Release documentation.

The reader interested in such results should request the following reports from the Science & Engineering Directorate:

- Division Report 2013 - 01(TR): Deliverable #1 – Analysis of risk rule effectiveness based on HS Code
- Division Report 2014-15(TR): Analysis of risk rule effectiveness based on HS Code

Using Other Standard Codes

In certain cases, the CBSA is provided a nicely codified value for the commodity being shipped to the country prior to arrival. This is mandated when a commodity that is to be transported to Canada is listed by the United Nations to be hazardous, and in such cases the trade chain partners must submit a properly classified UN Hazardous Goods Code. This allowed the Advanced Analytics team to compare the efficacy of the text-based fuzzy matching by using the UN Code instead of the HS Code provided on the Release documentation, and indeed the results were fairly consistent with the results described in the previous section.

What was found was that there were fewer codified rule hits, lending credence to the conclusion that Due to a technical detail in the way the Hazardous Goods rules are implemented it was difficult to ascertain how many of the shipments listing a Hazardous Goods Code also had the text-based rule fire against them;

Readers interested in this result should consult the following report:

- Division Report 2014 - 02(TR): Deliverable #4 – Analysis of risk rule effectiveness based on UN Hazardous Goods Code

Potential Solutions

Knowing that these sorts of problems existed, it was thought that another useful direction for the Advanced Analytics team's work would be to research cutting edge techniques that might serve to improve upon the current fuzzy matching process. The most promising direction this research took was to use data mining techniques to probabilistically translate the incoming text-based description into HS Codes that could then be compared to a list of potentially risky codes.

it was found that actionable information can be fairly accurately derived from the in-coming textual descriptions. More precisely, in the team's first attempt at tackling this problem, it was shown that a list of five candidate codes can be outputted with an confidence that the true HS Code would be in the list. This may sound a bit strange - typically one would like a more definitive answer - but when considering there are *a million* different possible classification at the six-digit level of the HS hierarchy, outputting a mere *five* candidates begins to look much more impressive.

This work was coded completely by the Advanced Analytics team in the Python programming language.

That being said, the next step in this analysis was an attempt to yield similar results by leveraging the procured data mining suite, IBM SPSS Modeller.

The original code (written in Python) had a distinct advantage in comparison to using the off-the-shelf Modeller software because the code could be finely tuned to the exact domain of the problem at hand. However, despite this potential lack of domain specific customization, it did turn out that the IBM suite was able to generate results roughly on par with that of the customized code. By outputting three candidate codes (note, the output list was kept this small to deal with memory issues) it was possible to have assurance that the correct commodity is in that list.

The interested reader should consult the following reports:

- Division Report 2013 - 03(TR): An Application of Naïve Bayes to the Classification of Commercial Cargo Declarations
- Division Report 2015-01(TR): Deliverable #1 – Use of IBM SPSS Modeller for Text Mining of Cargo Descriptions.

Conclusions

This was done by analysing various risk rules, and using two internationally recognized codings (both the Harmonized System Coding and the UN Hazardous Goods Coding) for commodity information as baselines for the study.

However, most importantly, the team was able to dispel the myth that the textual descriptions are of such poor quality that they are ill suited for use in an automated context. Indeed, two different methods were employed, one using technology already purchased by the Agency, in order to extract actionable information from the problematic descriptions.

It should also be noted that with more effort, refinements and improvements to this work might be possible to bolster the results described herein. With that in mind all of the work described in this memorandum should be looked at as pointing a potential direction for the Advanced Analytics team's focus.

Canada Border Services Agency | Agence des services frontaliers du Canada

eManifest Senior Project Advisory Committee Briefing

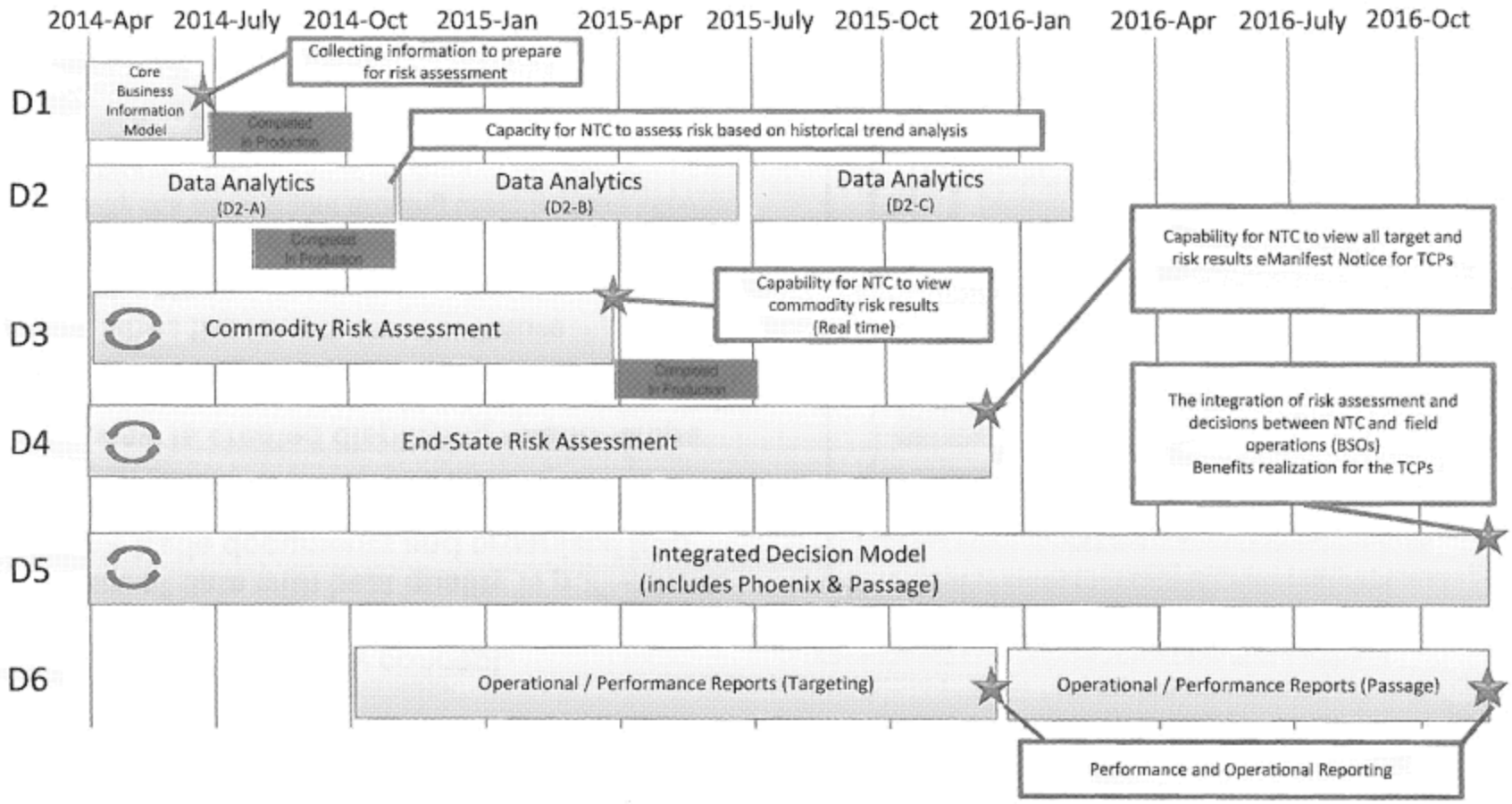
Project Update

June 15, 2015

Agenda

- eManifest Deployment Approach
  - Deployment 1 – Business Information Model
  - Deployment 2 – Data Analytics
  - Deployment 3 – Commodity Risk Indicators
  - Deployment 4 – End State Risk Assessment
  - Deployment 5 – Integrated Decision
  - Deployment 6 – Performance Reporting

eManifest Delivery Approach



Deployment 1: Entity Creation & Relationship

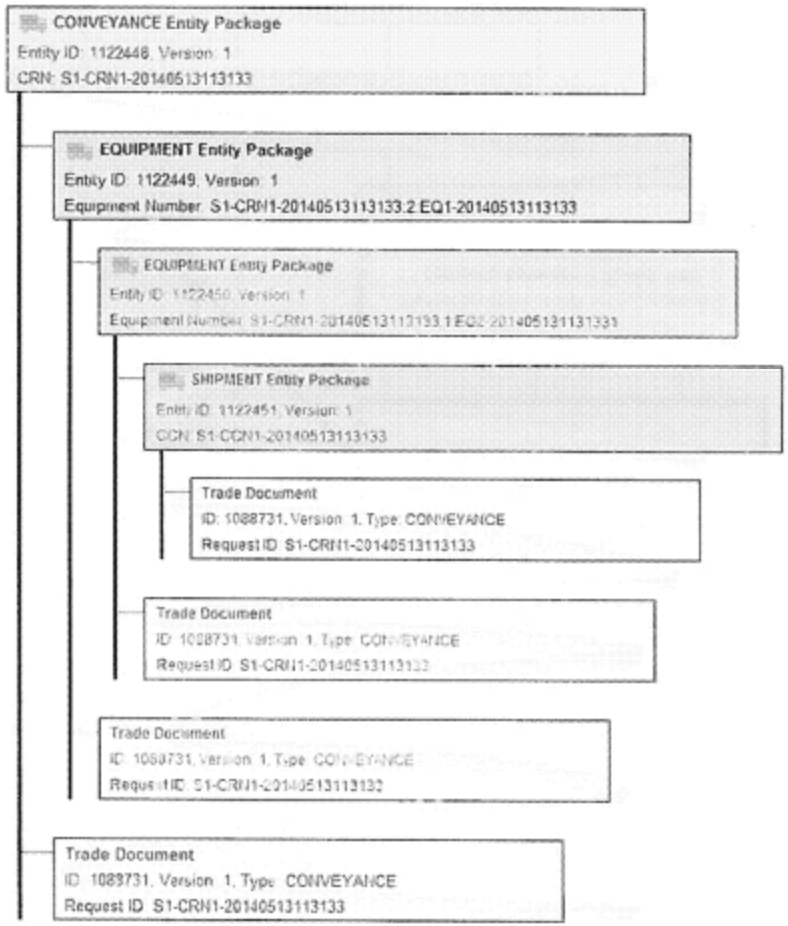
D1 introduces the entity concept.

Clusters of data with best quality is pulled from multiple trade documents and organized into entities – conveyance, equipment and shipment.

A trip view is created displaying relationships between conveyance, equipment and shipments.

Establishes the foundation for future deployments:

- Packages relevant data together to ensure timely and quality risk assessment allowing users to make informed decisions.
- Facilitating review of shipments by risk assessing and applying decisions at the entity level as opposed to individual documents.
- Risk results, decisions, referrals, and exam results are all associated to the entity enhancing the capability to perform analytics to improve future risk determination (i.e. "closing the loop").



Deployment 2: Data Analytics

Identification of shipments of interests based on Intelligence received from the National Targeting Centre (NTC).

Analytical tools that seek business pattern deviations based on known characteristics of a companies importing history.

Analytical tool to assist targeting operations in identifying illegal shipments of precursor chemicals.

Assesses pre-arrival commercial data and assists in identifying shipments of interest in the rail environment.

Deployment 3 – Commodity Risk Assessment

- Deployment 3 is the foundation for commercial risk assessment;
- D3 provides the NTC with the opportunity to use the new capability in an operational environment and to provide feedback for future (D4/D5) deployments;
- D3 delivers the ability to perform automated risk assessment for shipment entities in all four modes using Commodity based risk indicators;
- D3 will help shape operational processes associated with using a business rules engine to create, modify and/or turn off commercial risk rules based on performance or emerging threat categories; and
- Introduce the new commercial targeting system (Phoenix) for searching/viewing risk results.

Deployment 4

"End-State Risk Assessment"

Scope

- Complete Automated Risk Assessment (all risk rules are executing and viewable)
- Implementation of initial eManifest new notices for Trade Chain Partners (TCPs)
- Implementation of the Single Window trade document (Integrated Import Declaration) as a release option
- Resolved identities of TCPs using Master Data Management
- Implementation of a risk rules simulation environment

Business Outcomes

- Capability for the NTC to view targets and all risk results (Real Time) in all modes
- Capability to assess the operational impact of implementing new risk rules (using simulation)
- The new notices provide desirable functionality to help improve communication between CBSA and its clients as well as business-to-business communication.
- Validation of the Risk Assessment Model (identification of low and high risk entities)
- Validation that the planned targeting work force can handle the volume

Target Production Date: December 2015

Deployment 5

"Integrated Decision Model"

Scope

- Integrated decisions and referrals (Risk Assessment, Passage and Single Window Initiative)
- Capture of examination results by front line operations
- End-state notices via Electronic Data Interchange and eManifest Portal
- Introduction of Advance Trade Data (ATD) from Importers
- Implementation of end state eManifest trade document submission

Business Outcomes

- Complete integration of risk assessment and passage decisions between NTC and field operations Border Services Officers (BSOs)
- Enhance Program integrity through "closing the loop" on examination results
- Advance Trade Data (ATD) in all modes supports Targeting Program – provides clarity on what commodities are being imported by whom
- Fully integrated commercial processing system and application, includes SWI
- New Documents and Notices available to external clients
- The eManifest system becomes the new system of record
- Full Benefits Realized for TCPs (Manifest Forward, Streamlined Border Processing)

Target Production Date: December 2016

Deployment 6

« *Performance Reporting* »

- 36 Reports
  - Split in 2 Releases
    - D6A – aligns with D4B (4 Reports) – implementation post-D4B
    - D6B – aligns with D5B (32 Reports) – implementation post-D5
- Audit – DSO Audit capabilities for COGNOS reports
- COGNOS Workspace upgrades
- Required hardware uplift to COGNOS environments

Next Steps

- Continue with project transition and support of Deployment 3.
  - Evaluate/tune the commodity risk rules using live production data;
  - Conduit for regular D3 feedback that will guide future developments; and
  - Manage the D3 roll-out (training/support).
- Continue development and testing of remaining project deployments.

Canada Border Services Agency | Agence des services frontaliers du Canada

eManifest Senior Project Advisory Committee

Project Update

December 23, 2015
Commercial Projects Directorate
eManifest Division

PROTECTION · SERVICE · INTEGRITY

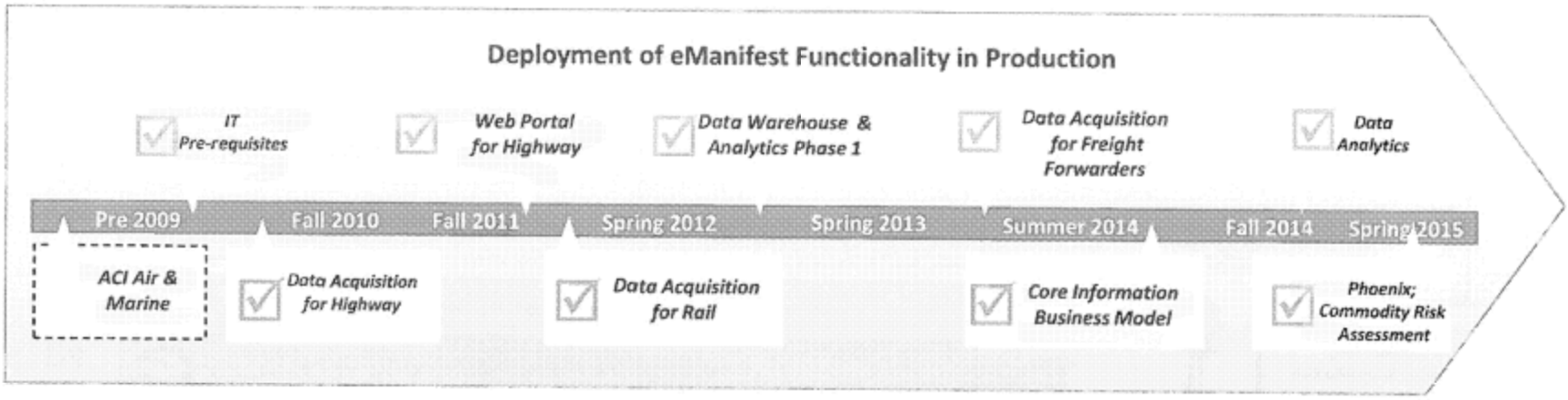
Canada

Overview

- eManifest modernizes and enhances commercial processes and screening of Canada-bound goods by improving the CBSA's ability to detect shipments that pose a high or unknown risk, while facilitating the movement of low-risk shipments across the border prior to arrival.

Data Acquisition	Risk Assessment	Integrated Decision Model (Passage)	Business Intelligence
			
Mandating the acquisition of electronic pre-arrival data	Automating the risk assessment of all data in advance of arrival at the border	Complete integration of risk assessment between the NTC and Field Operations BSOs	Introducing integrated Business Intelligence systems

eManifest Accomplishments to Date



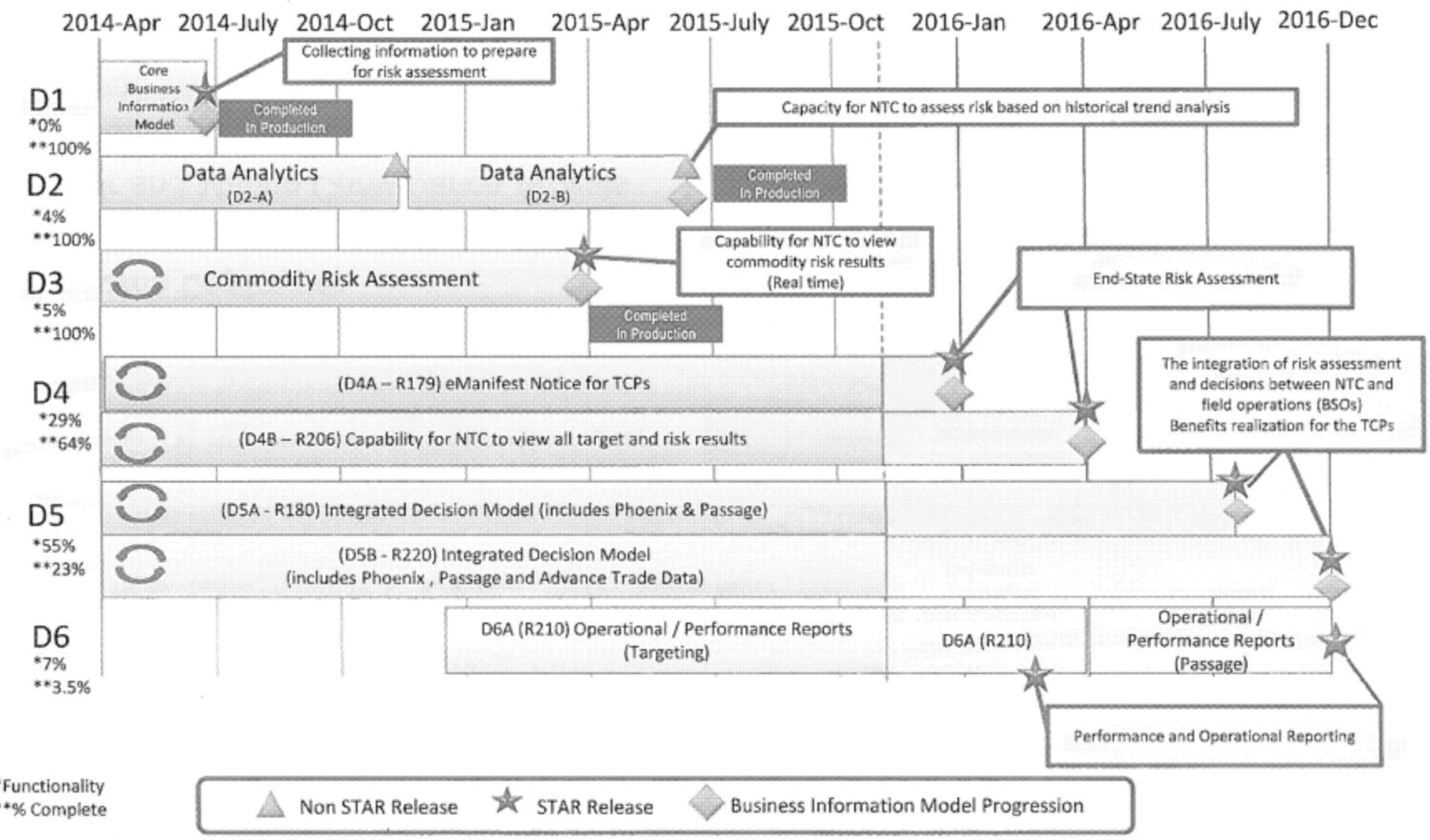
Systems Deployed:

- ✓ eManifest Portal
- ✓ Air and Marine Conveyance Arrivals
- ✓ Manifest Forward
- ✓ Data Warehouse
- ✓ Core Information Business Model
- ✓ Data Analytics
- ✓ Phoenix

Reporting:

- ✓ Highway Reporting
- ✓ Rail Reporting
- ✓ Freight Forwarder Reporting

Schedule



Deployment 2

- The business case for D2 has been achieved with the successful completion of two releases:
  - D2A: The project implemented a subset of the Commercial Mining Mart on the new EDW Appliance with supporting data analytics to demonstrate the business value of moving to a 24 hour refresh rate. Production Date: October 22, 2014
  - D2B: The project augmented the 24 hour refresh of the D2A data sources with Business Information Model (BIM) and Commodity Risk Results data sources. Production Date: June 29, 2015

Deployment 4A – Notices (January 2016)

Winter 2016, the CBSA will introduce new and enhanced notification systems to increase automation of pre and post-arrival notices to clients on commercial movements.

- A subset of new eManifest notices, for EDI clients, will advise on the completeness of advance data submitted to the CBSA and on the arrival and release statuses of shipments.

- The new eManifest notices will be based on the overall shipment status.
  - Current notices are based on individual documents submitted by clients; Release Notification System (RNS) messages are based on the status of the release only.
  - The new eManifest notices will be based on the overall shipment status and notices will be transmitted against individual documents within a shipment whenever the shipment status changes.

- Adoption of the new subset of eManifest notices will be voluntary to provide EDI clients the opportunity to transition to the end-state eManifest notices.

eManifest Notices

Notice	Importer / Broker	Carrier / Freight Forwarder	Warehouse Operator
Completeness Notices			
• Matched	✓	✓	
• Not Matched	✓	✓	
• Cargo Complete		✓	
• Document Package Complete		✓	
Disposition Notices			
• ***Arrival Notices***			
▪ Reported		✓	✓
▪ Arrived		✓	✓
• ***Status Notices***			
▪ Deconsolidation		✓	✓
▪ Document Not on File		✓	
▪ Authorized to Deliver	✓	✓	✓
▪ Released	✓	✓	✓
▪ Held (Basic)	✓	✓	✓

eManifest Benefits

For the Government of Canada

Category	Benefit Description
Security	• Entity Model – Pre-arrival trade data from multiple sources is compiled into "entities" that provide a complete view of the shipment. Risk assessment happens at the entity level rather than on individual trade documents. • Automated Risk Assessment – Commercial entities, across all modes, are risk assessed prior to arrival. High and unknown risk entities are flagged to the targeting officers. Low risk shipments directed to move inland or released by the system. Provides the Agency with the ability to respond to threats in a more timely manner – hours versus weeks. • Integrated Targeting Model – Risk assessment for admissibility and release decisions in all four modes is done by targeting officers at the NTC. CBSA is offered better consistency, as all modes are aligned. • Business Intelligence – Risk assessment results and examinations are used to "close the loop", revise risk assessment rules and continually improve the risk assessment process. • Electronic data will inform other programs to enhance their risk assessment.
Service	• Integrated Targeting Model – Risk assessment for admissibility and release decisions in all four modes is done by targeting officers at the NTC. CBSA is offered better consistency, as all modes are aligned. • Business Intelligence – Risk assessment results and examinations are used to "close the loop", revise risk assessment rules and continually improve the risk assessment process. • Electronic data will inform other programs to enhance their risk assessment.
Savings	• Integrated Targeting Model – Risk assessment for admissibility and release decisions in all four modes is done by targeting officers at the NTC. CBSA is offered better consistency, as all modes are aligned. • Business Intelligence – Risk assessment results and examinations are used to "close the loop", revise risk assessment rules and continually improve the risk assessment process. • Electronic data will inform other programs to enhance their risk assessment.

eManifest Benefits (con't)

For trade and Canadians

Category	Benefit Description
Service	• Electronic Data Interchange – Trade Chain Partners electronically transmit advanced commercial information, which can be used in communication with other trade chain partners. • Predictability – Risk assessment prior to arrival provides predictability for trade. • Consistency – National integrated targeting provides trade with consistency, as all modes in the Commercial process are aligned. • Facilitation – focus on high risk shipments for targeting and inspection facilitates the legitimate flow of low-risk trade. • Ability to cleanse data reduces costly requests for information (RFI) to Trade.

ANNEX

Cost

Prior to Rebaselining: $295.1M (including EBP fully loaded)		
Deployment (Sub-Deployments)	**Budget**	**Completed on Time and on Budget**
Core Business Information Model *	$10.3M	✓
Data Analytics (A)	$1.1M	✓
Data Analytics (B)	$3.4M	✓
Commodity Risk Assessment	$21.8M	✓
End State Risk Assessment	$34.9M	Ongoing
Integrated Decision Model	$42.4M	Ongoing
Operational / Performance Reports	$16.3M	Ongoing
Total for All Deployments	**$119.9M**	

Dashboard

Executive Project Dashboard

Organization: Canada Border Services Agency
Project: eManifest

Project Phase: Execution
Report as of: 2015-Oct-31

Project Complexity and Risk Rating: Level 4 – Transformational
Next Project Gate: 6 – Construction Complete and Deployment Readiness

Project Health Y

Executive Summary

Project Sponsor: Maurice Chenier, Martin Belduc

Current Period: Deployment 4A introduces a Notification system that clients will use to track their importations. Software testing was extended to meet quality management objectives. Revised target date for software testing completion is December 15, 2015.
Forecast: Formalize the Deployment 4A revised production date through governance.
Overall Status: The project overall health is yellow. Remaining risks and issues are being mitigated.

Business Outcomes

- Enhance CBSA capacity to provide a pre-arrival risk determination prior to the arrival of goods in Canada
- Provide CBSA with the ability to conduct more effective enforcement activities
- Enable CBSA, PGAs & TCPs to evolve toward an automated eCommerce importation process in line with international standards.

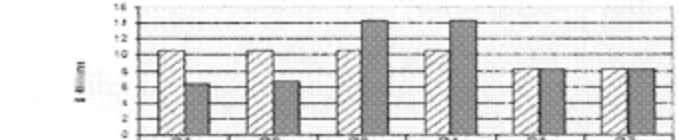
Project Cost: Approved Budget : $415.1 million
Actual/Forecast Cost : $415.1 million G

YYYY/YY	Prior	2012-13	2013-14	2014-15	2015-16	2016-17		Total
Approved	$378.0	$37.1	$0.0	$0.0	$0.0	$0.0		$415.1
Actual/Forecast	$228.5	$33.5	$36.9	$33.5	$41.6	$32.9		$415.1

Development Only

Cost	G
Schedule	G
Scope	G
Risk	Y
Issues	Y

Financial Summary (six quarters)



	Q1 - 15/16	Q2 - 15/16	Q3 - 15/16	Q4 - 15/16	Q1 - 16/17	Q2 - 16/17
Approved	$10.39	$10.39	$10.39	$10.39	$8.22	$8.22
Actual / Forecast	$6.41	$6.69	$14.24	$14.24	$8.25	$8.35

Project Schedule

Project Launch Date: 2006-Oct-05 G

Key Milestones/Deliverables	Approved Completion Date	Forecast Completion Date	Variance (months)
Deployment 4A (R173) : Risk Assessment (Notices)	31-Dec-2015	19-Dec-2015	0.39
(R173) Sub Activity : User Acceptance Testing	9-Dec-2015	4-Dec-2015	0.16
ODM Re-Platform (R219)	30-Nov-2015	30-Nov-2015	0.00
Deployment 4B (R206), End State Risk Assessment	31-Dec-2015	2-Apr-2016	-3.00
(R206) Sub Activity : Development	15-Jan-2016	15-Jan-2016	0.00
Deployment 6A (R210) : Operational / Performance Reports	3-Dec-2015	2-Apr-2016	-3.90
(R210) Gating - Business Requirements Review (BRR)	24-Nov-2015	24-Nov-2015	0.00
Deployment 5A (R180) : Integrated Decision Model (Phoenix/Passage Applicatio	16-Dec-2016	31-Aug-2016	3.45
(R180) Sub Activity : Development	30-Mar-2016	30-Mar-2016	0.00
(R180) Sub Activity : User Acceptance Testing	17-Aug-2016	17-Aug-2016	0.00
Deployment 5B (R220): Integrated Decision Model (Phoenix/Passage Applicatio	16-Dec-2016	16-Dec-2016	0.00
(R220) Sub Activity : Design and Training Product Development	16-Dec-2016	16-Dec-2016	0.00
Deployment 6B: Operational / Performance Reports	16-Dec-2016	16-Dec-2016	0.00
Project Completion	**2017-Mar-31**	**2017-Mar-31**	**0**

Project Risk

Y

Top Risks

Project Scope / Requests for Change (RFC)

G

	Approved RFC	Estimated $ Effort Days	Estimated Cost	Impact On Project
This Period				
To Date	42	555	$1.547M	Cost included in rebaseline. Zero impact to project

Project Issues

Y

Issues	New	Open	Closed
This Period	0	4	0
Total		4	350

Open Issues	Resolution Plan	Due Date	Accountability
(IL-119) Procurement - RFP (Professional Services)	Presently working with CBSA contracting and PSPC/SPI to finalize contracting documentation in order to post the RFP competitively. Project team have engaged senior management at both the CBSA and PSPC/SSC to discuss mitigation options. Options have been identified to eliminate impact.	2015 Nov 30	B Radoz, DG Commercial Projects
IL-122 (Testing Strategy)	Testing strategy has been developed and endorsed (October 16, 2015) by all testing teams. Implementation of testing strategy has commenced. Activities include, augmenting resources, review of all deployment schedules to ensure that the adequate testing time has been allocated. Analysis is underway to form a single unified testing team for development testing community.	2015-Dec-31	B Radoz, DG Commercial Projects
(IL-121) User Acceptance Testing (UAT) for D4A was delayed due to bugs (blockers) uncovered during system integration testing	The team is currently completing a broader review of testing activities/requirements and impacts to timelines. Prioritization exercise of test cases is underway for D4A (Notices) this will allow the testing team to focus on the mandatory business requirements.	2015-Nov-30	B Radoz, DG Commercial Projects

Deployment 2

"Data Analytics"

Scope

- Data Analytics capability to assist intelligence officers / targeting teams in the National Targeting Centre (NTC) in the mining of the existing and historical trade data

Business Outcomes

- Capability for the NTC to assess risk based on historical trend analysis (e.g. anomalies in a companies Business Profile – pattern and trend deviation)
- Identification of 'candidate' risk indicators (e.g. use analytics to develop new rules based on vessel routing patterns, container delivery address)
- Modification of existing indicators based on analysis / outcomes and new data feeds

Target Production Date:

- ✓ Deployment 2A: Implemented in Production - October 2014
- ✓ Deployment 2B: June 2015

Deployment 3

"Commodity Risk Assessment"

Scope

- Start of Automated Risk Assessment
- Introduction of the risk results User Interface (UI) that supports the viewing of shipments
- Provide the ability to view and modify High Risk Commodity rules
- Implementation of High Risk Commodity rules to support Automated Risk Assessment of shipments

Business Outcomes

- Capability for the NTC to view High Risk Commodity risk results (Real Time) in all modes;
- Supports the ability to target or interdict high risk shipments using legacy commercial systems
- Ability for the Program to assess the performance of High Risk Commodity rules in new system vs. legacy system
- Validating and improving the Automated Risk Assessment results

Target Production Date:

✓ March 2015

Deployment 4

"End-State Risk Assessment"

Scope

- Complete Automated Risk Assessment (all risk rules are executing and viewable)
- Implementation of initial eManifest new notices for Trade Chain Partners (TCPs)
- Implementation of the Single Window trade document (Integrated Import Declaration) as a release option
- Resolved identities of TCPs using Master Data Management
- Implementation of a risk rules simulation environment

Business Outcomes

- Capability for the NTC to view targets and all risk results (Real Time) in all modes
- Capability to assess the operational impact of implementing new risk rules (using simulation)
- The new notices provide desirable functionality to help improve communication between CBSA and its clients as well as business-to-business communication.
- Validation of the Risk Assessment Model (identification of low and high risk entities)
- Validation that the planned targeting work force can handle the volume

D4A Target Production Date: January 30-31, 2016
D4B Target Production Date: April 2-3, 2016

Deployment 5

"Integrated Decision Model"

Scope

- Integrated decisions and referrals (Risk Assessment, Passage and Single Window Initiative)
- Capture of examination results by front line operations
- End-state notices via Electronic Data Interchange and eManifest Portal
- Introduction of Advance Trade Data (ATD) from Importers
- Implementation of end state eManifest trade document submission

Business Outcomes

- Complete integration of risk assessment and passage decisions between NTC and field operations Border Services Officers (BSOs)
- Enhance Program integrity through "closing the loop" on examination results
- Advance Trade Data (ATD) in all modes supports Targeting Program – provides clarity on what commodities are being imported by whom
- Fully integrated commercial processing system and application, includes SWI
- New Documents and Notices available to external clients
- The eManifest system becomes the new system of record
- Full Benefits Realized for TCPs (Manifest Forward, Streamlined Border Processing)

D5A Target Production Date: August 2016
D5B Target Production Date: December 2016

Deployment 6

"Operational and Performance Reports"

Scope

- Risk Assessment: Operational and Management Reports
- Passage: Operational and Management Reports

Business Outcomes

- Program Performance and Operational Reporting
- Increased Decision Support for Programs

Target Production Date:

- Deployment 6A (aligns with D4B production): April 2016
- Deployment 6B (aligns with D5B Production): December 2016

Canada Border Services Agency | Agence des services frontaliers du Canada

eManifest Project Update for Commercial Programs Directorate

Commercial Projects Directorate
May, 2015

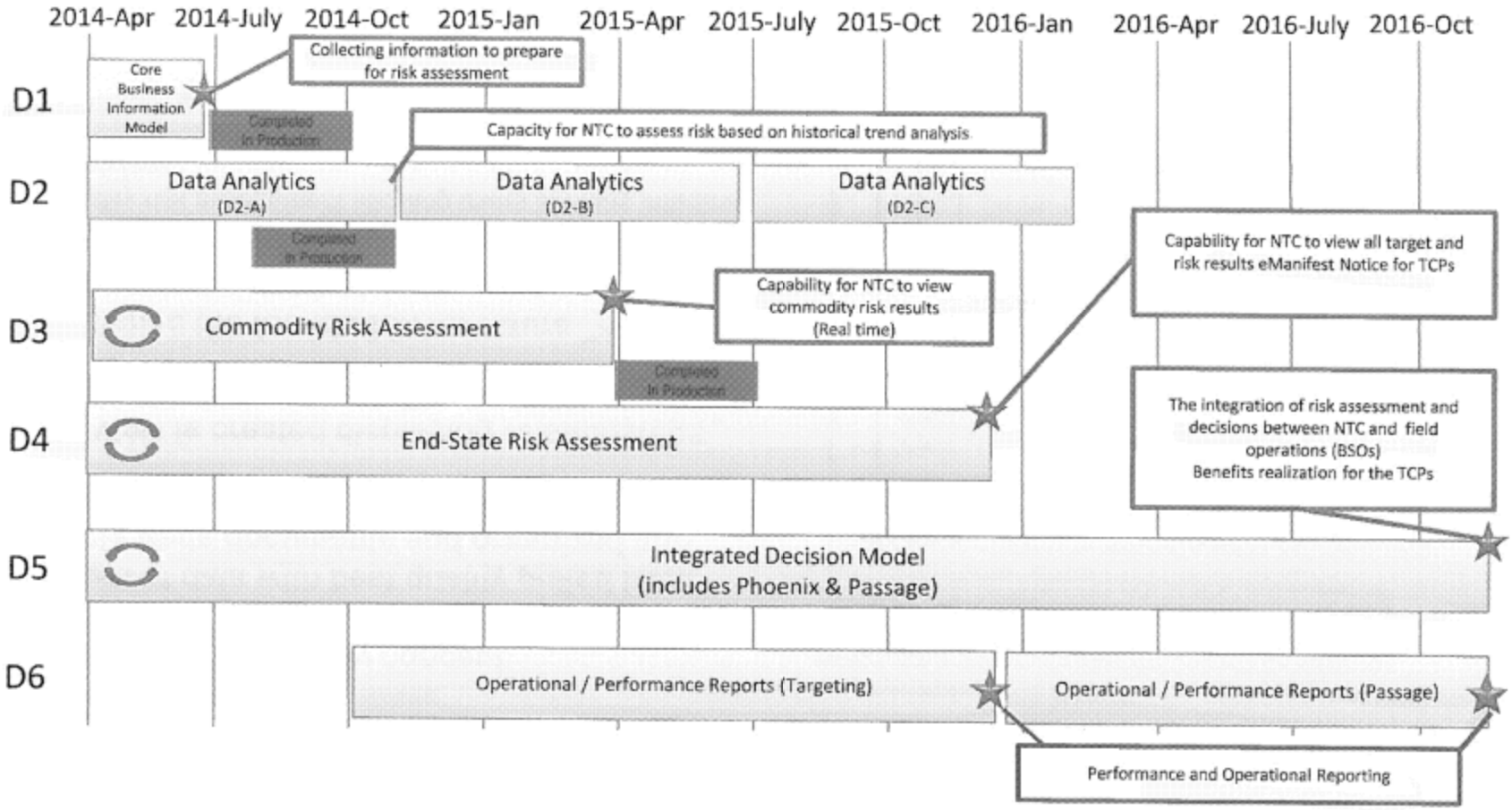
PROTECTION • SERVICE • INTEGRITY

Canada

Agenda

- eManifest Deployment Overview
  - Deployment 1 – Business Information Model
  - Deployment 2 – Data Analytics
  - Deployment 3 – Commodity Risk Indicators
  - Deployment 4 – End-State Risk Assessment
  - Deployment 5 – Integrated Decision Model
  - Deployment 6 – Operational and Performance Reports

eManifest Delivery Approach



Deployment 1: Entity Creation & Relationship

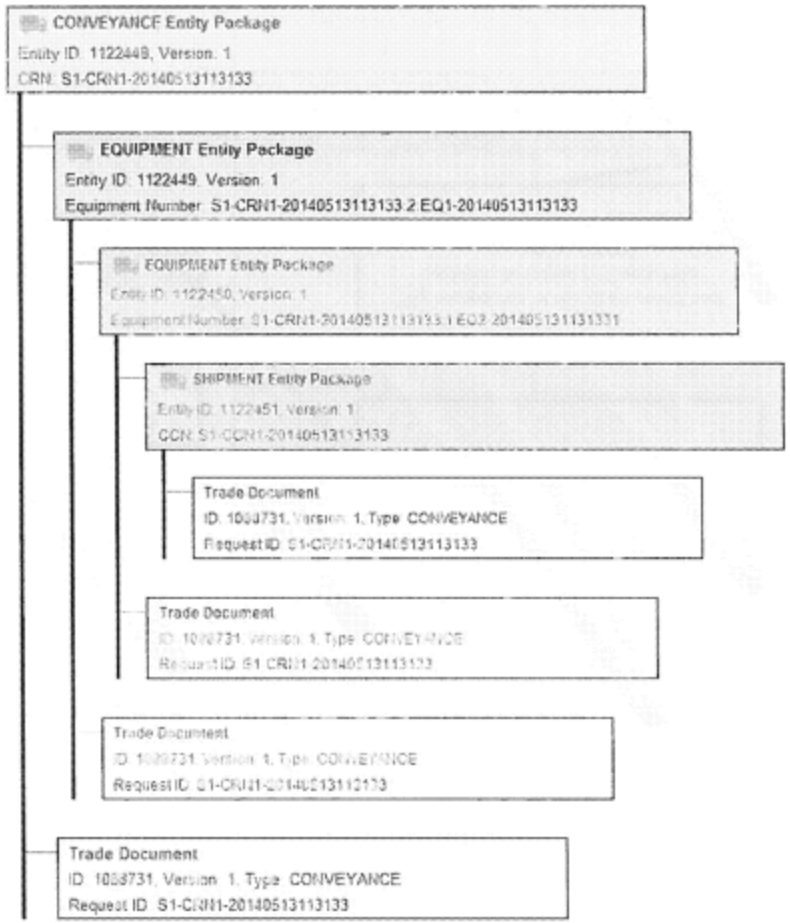
D1 introduced the entity concept.

Clusters of data with best quality pulled from multiple trade documents and organized into entities – conveyance, equipment and shipment.

A trip view is created displaying relationships between conveyance, equipment and shipments.

Established the foundation for future deployments:

- Packages relevant data together to ensure timely and quality risk assessment allowing users to make informed decisions.
- Facilitating review of shipments by risk assessing and applying decisions at the entity level as opposed to individual documents.
- Risk results, decisions, referrals, and exam results are all associated to the entity enhancing the capability to perform analytics to improve future risk determination (i.e. "closing the loop").



Deployment 2: Data Analytics

Identification of shipments of interests based on Intelligence received from the National Targeting Centre (NTC).

Analytical tools that seek business pattern deviations based on known characteristics of a companies importing history.

Analytical tool to assist targeting operations in identifying illegal shipments of precursor chemicals.

Assesses pre-arrival commercial data and assists in identifying shipments of interest in the rail environment.

Deployment 3

"Commodity Risk Assessment"

- Deployment 3 is the foundation for commercial risk assessment;
- D3 provides the NTC with the opportunity to use the new capability in an operational environment and to provide feedback for future (D4/D5) deployments;
- D3 delivers the ability to perform automated risk assessment for shipment entities in all four modes using Commodity based risk indicators;
- D3 will help shape operational processes associated with using a business rules engine to create, modify and/or turn off commercial risk rules based on performance or emerging threat categories; and
- Introduce the new commercial targeting system (Phoenix) for searching/viewing risk results.

Deployment 4 (D4A-R179 & D4B-TBD)

"End-State Risk Assessment"

- Deployment 4 is being delivered in two releases.
- The below tables represent what components are being delivered in D4A and D4B.

D4A – Data Acquisition

- Notices
- Data Transition
- IID Consumption
- Identity Resolution
- Risk Assessment
- Audit

D4B – Risk Assessment

- Data Acquisition
- Data Transition
- Identity Resolution
- Risk Assessment
- Audit

ODM re-platform

Deployment 5 (R180)

"Integrated Decision Model"

- D5 will introduce the new business information model which will change how commercial information is stored and decisions are made
- New Commercial border processing application (Passage) that will be used by BSOs to process commercial shipments as they enter Canada
- Full integration of Passage and new risk assessment systems
- End-state notices via Electronic Data Interchange and eManifest Portal
- Introduction of Advance Trade Data (ATD) from Importers
- Implementation of end state eManifest trade document submission
- Enhanced program integrity through "closing the loop" on examination results
- eManifest systems becomes the new systems of record
- Full benefits realized for TCPs (Manifest Forward, Streamlined Border Processing) and internal stakeholders (streamlined examination results, better application usability, etc.)

Canada Border Services Agency | Agence des services frontaliers du Canada

eManifest Deployment 2

Presentation to Project Portfolio Advisory Committee (PPAC)

September 03, 2015

eManifest Division (Mike Leahy)

PROTECTION • SERVICE • INTEGRITY

Canada

Objective

- eManifest is seeking approval that:
  - The delivery of Deployment 2 (D2) project objectives have been achieved as per the business case, and
  - Future data analytics enhancements will be covered under maintenance funding.

Background

- The business case, as outlined in the TB Submission, included a deliverable for Advanced Data Analytics for the amalgamation of the advanced analytics platform to include all modes.

- The business case for D2 has been achieved with the successful completion of two releases:

  - D2A: The project implemented a subset of the Commercial Mining Mart on the new EDW Appliance with supporting data analytics to demonstrate the business value of moving to a 24 hour refresh rate. Production Date: October 22, 2014

  - D2B: The project augmented the 24 hour refresh of the D2A data sources with Business Information Model (BIM) and Commodity Risk Results data sources. Production Date: June 29, 2015

Future Maintenance Activities

- Maintenance Release
  - Will augment the refresh of the D2A & D2B data sources with additional Business Information Model, Master Data Management, Risk Results and Target Results data sources.
  - Will be implemented as part of Service Management under Commercial Projects Directorate (CPD) and adhere to SLMF.

Next Steps

- Communicate D2 close-out decision to all stakeholders and delivery teams.
- Future enhancements will be aligned with maintenance funding.

Canada Border Services Agency | Agence des services frontaliers du Canada

eManifest Project Recovery

Delivery Approach

Senior Project Advisory Committee (SPAC)

July 16, 2014

Background

- In February the CBSA began to perform a review of the eManifest project and a root cause analysis began to identify risks, mitigation strategies and develop a Delivery Framework for the to recover the project with full business benefits.
- In March three Options for moving forward were identified and an Option Analysis was undertaken. The following three Options for moving forward were identified:

Option 2 – Hybrid Delivery

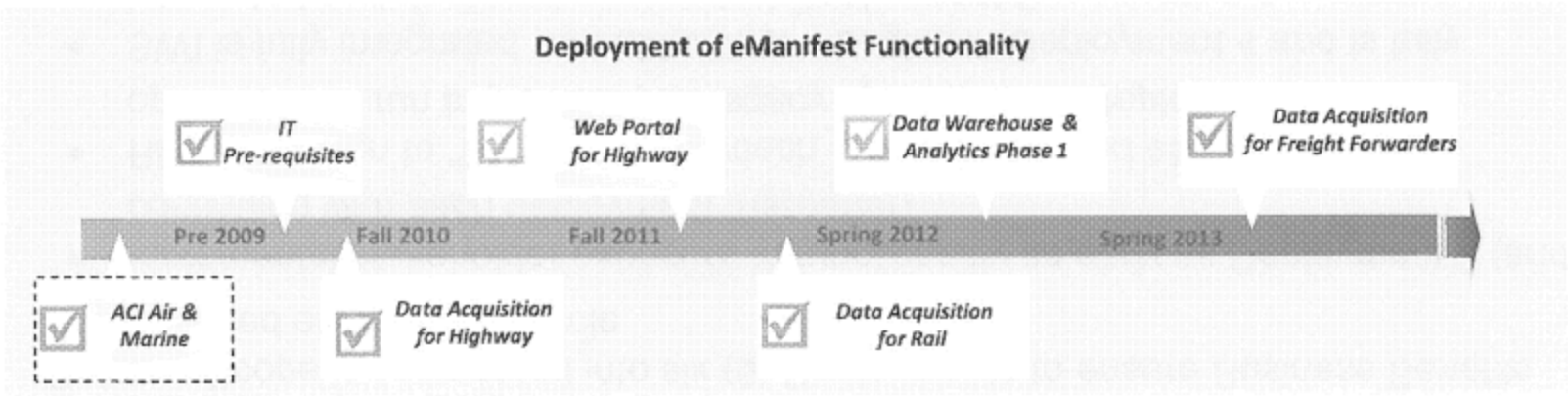
  - This option introduces a new delivery model designed to deliver business benefits early by augmenting the project team competency and capacity with different forms of vendor relationships

- In April the Executive Committee (EC) approved moving forward with Option 2 and the project team proceeded with the development of the detailed planning, risk mitigation and third party endorsement activities.
- In June, the new delivery approach was approved by EC.

New Delivery Approach

- ✓ The new delivery approach is supported by CBSA resources and is augmented with external support to do parallel work that will generate early, tangible benefits to confirm development milestones are being met and provide support for front-line operations.
- ✓ The revised plan will provide additional time to train CBSA staff and conduct outreach to external stakeholders to ensure a successful implementation.
- ✓ The plan addresses external clients' requests for more time to make changes to their internal business processes and systems in order to adapt to the new requirements.
- ✓ Delivery Scope is segmented into six (6) Deployments to ensure Business Benefits are delivered earlier in the cycle
  - Incremental stakeholder exposure to functionality as early as Deployment 2 (and continuing with each deployment thereafter)
  - Risk mitigation to "Big Bang" approach - newly deployed eManifest system operations to run in parallel with legacy systems until Deployment 5
  - SWI is fully integrated with eManifest – begins in Deployment 4 and is fully integrated and deployed as a part of Deployment 5
  - Crew and electronic re-manifest requirements have been removed from scope and will be delivered as a part of the larger CBSA border modernization program

eManifest Key Accomplishments to Date



Systems Deployed:

- ✓ Highway Cargo and Conveyance Reporting
- ✓ Rail Reporting
- ✓ eManifest Portal
- ✓ Freight Forwarder Reporting
- ✓ Air and Marine Conveyance Arrivals
- ✓ Manifest Forward
- ✓ Data Warehouse

Implementation Highlights:

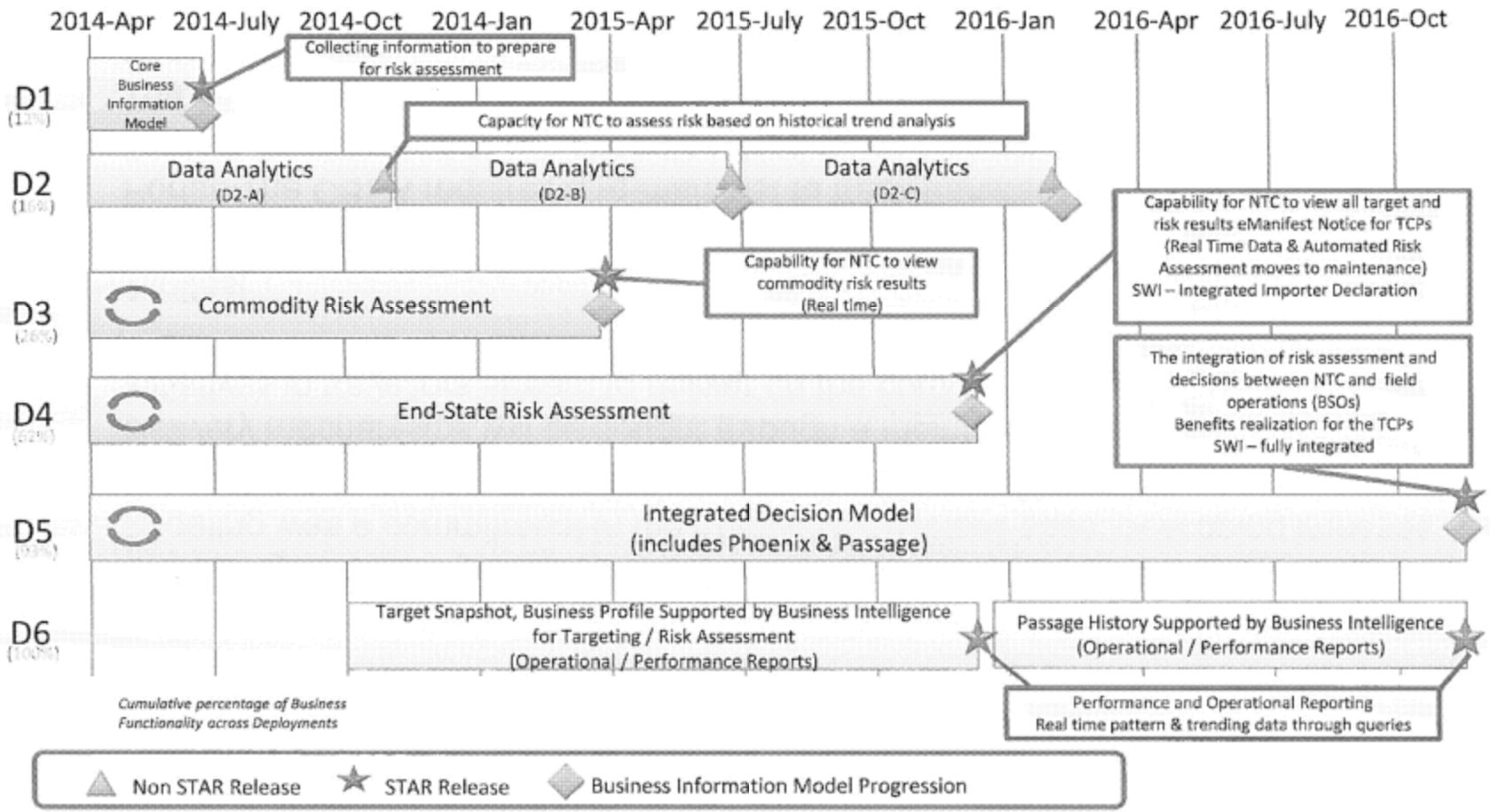
- ✓ Over 11,000 highway carriers now engaged with eManifest
- ✓ 97% of Top 500 Carriers engaged
- ✓ Regulations progressing
- ✓ Multiple rounds of training for BSOs
- ✓ Regional network in place
- ✓ Established a program alignment structure to resolve outstanding issues
- ✓ Webinars, web content, presentations

De-scoping impacts

- Crew
  - The ability to collect pre-arrival crew information for risk assessment and targeting was a commitment in the eManifest business case; descoping imposes the continued need for paper based manual review and targeting.
  - Delivery mechanisms will be sought through a maintenance release and/or joint initiatives in travellers stream to deliver on the commitment.

- Re-Manifest
  - Functionality for carriers to submit re-manifests electronically through EDI or the Portal. The CBSA requires a re-manifest to track transfers in liability between carriers or when goods are changing destination / warehouse location.
  - The agency is reviewing the business value of re-manifest in light of agency transformational initiatives; stakeholders support this review as it offers business process efficiencies.
  - Possible opportunity in changing requirement through the Cargo Control and Sufferance Warehouse Modernization Initiative

Delivery Approach Designed to Generate Early Benefits



Deployment 1

"Core Information Business Model"

Scope

- Electronic Data Capture
- Laying the foundation to prepare the trade documents to support risk assessment (building entity relationships)
  - Description of the High Risk Commodity identification supported by Language Ware.
  - Address data preparation supported by Quality Stage–Address Verification Interface (QS-AVI)

Business Outcomes

- Validation that the core business design is sound
- Validation that the implemented COTS products provide the expected capability (e.g. ability to resolve an address – supports risk rules that look for commercial shipments destined to residential address, determine what type of commodity is being reported on the manifest)

✓ Implemented in Production – June 2014

Deployment 2

"Data Analytics"

Scope

- Data Analytics capability to assist intelligence officers / targeting teams in the National Targeting Centre (NTC) in the mining of the existing and historical trade data

Business Outcomes

- Capability for the NTC to assess risk based on historical trend analysis (e.g. anomalies in a companies Business Profile – pattern and trend deviation)
- Identification of 'candidate' risk indicators (e.g. use analytics to develop new rules based on vessel routing patterns, container delivery address)
- Modification of existing indicators based on analysis / outcomes and new data feeds

- **Target Production Date:**
  - Deployment 2A: October 2014
  - Deployment 2B: June 2015
  - Deployment 2C: February 2016

Deployment 3

"Commodity Risk Assessment"

Scope

- Start of Automated Risk Assessment
- Introduction of the risk results User Interface (UI) that supports the viewing of shipments
- Provide the ability to view and modify High Risk Commodity rules
- Implementation of High Risk Commodity rules to support Automated Risk Assessment of shipments

Business Outcomes

- Capability for the NTC to view High Risk Commodity risk results (Real Time) in all modes;
- Supports the ability to target or interdict high risk shipments using legacy commercial systems
- Ability for the Program to assess the performance of High Risk Commodity rules in new system vs. legacy system
- Validating and improving the Automated Risk Assessment results
- **Target Production Date:** March 2015

Deployment 4

"End-State Risk Assessment"

Scope

- Complete Automated Risk Assessment (all risk rules are executing and viewable)
- Implementation of initial eManifest new notices for Trade Chain Partners (TCPs)
- Implementation of the Single Window trade document (Integrated Import Declaration) as a release option
- Resolved identities of TCPs using Master Data Management
- Implementation of a risk rules simulation environment

Business Outcomes

- Capability for the NTC to view targets and all risk results (Real Time) in all modes
- Capability to assess the operational impact of implementing new risk rules (using simulation)
- The new notices provide desirable functionality to help improve communication between CBSA and its clients as well as business-to-business communication.
- Validation of the Risk Assessment Model (identification of low and high risk entities)
- Validation that the planned targeting work force can handle the volume

Target Production Date: December 2015

Deployment 5

"Integrated Decision Model"

Scope

- Integrated decisions and referrals (Risk Assessment, Passage and Single Window Initiative)
- Capture of examination results by front line operations
- End-state notices via Electronic Data Interchange and eManifest Portal
- Introduction of Advance Trade Data (ATD) from Importers
- Implementation of end state eManifest trade document submission

Business Outcomes

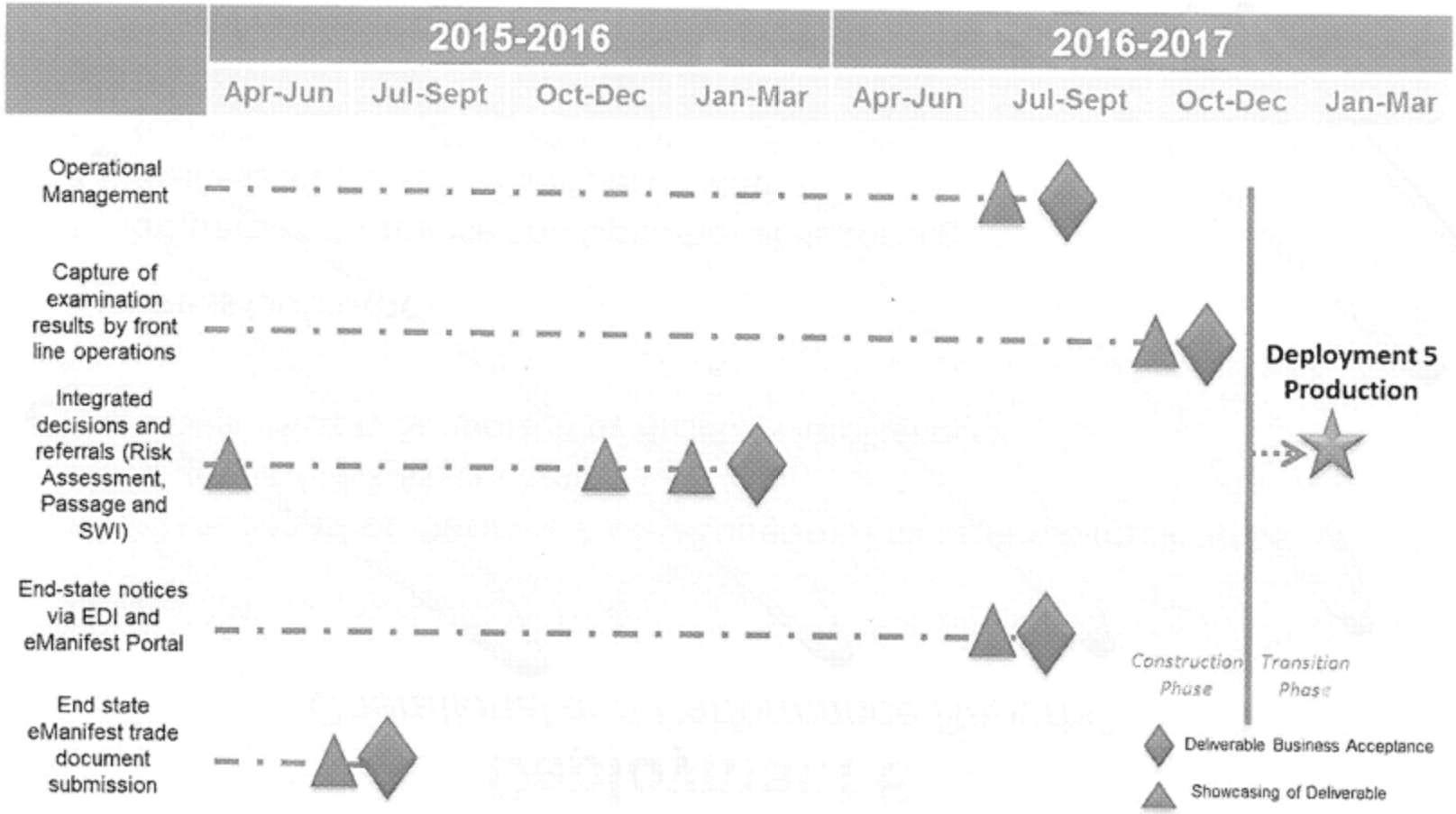
- Complete integration of risk assessment and passage decisions between NTC and field operations Border Services Officers (BSOs)
- Enhance Program integrity through "closing the loop" on examination results
- Advance Trade Data (ATD) in all modes supports Targeting Program – provides clarity on what commodities are being imported by whom
- Fully integrated commercial processing system and application, includes SWI
- New Documents and Notices available to external clients
- The eManifest system becomes the new system of record
- Full Benefits Realized for TCPs (Manifest Forward, Streamlined Border Processing)

- **Target Production Date:** December 2016

D5 Business Outcome Realization

- D5 rollout is structured around the completion of components such as risk assessment and data acquisition to create efficiencies in delivery of the passage requirements
- D5 is broken down into five deliverables that integrate together to deliver the full benefits realization
- Importer end-state eManifest trade documents will be delivered in D5 to align with the new system and avoid redundant costly work on ACROSS
- As a result, the integrated decision model is dependant on completion of the risk assessment component aligning risk assessment and passage as an integrated system of record

Deployment 5 - Deliverables & Timelines



Deployment 6

"Operational and Performance Reports"

Scope

- Target Snapshot, Business Profile Supported by Business Intelligence for Targeting / Risk Assessment
- Passage History Supported by Business Intelligence

Business Outcomes

- Program Performance and Operational Reporting
- Business Intelligence – Self Serve Model
- Real time pattern & trending data through queries

- **Target Production Date:**
  - Deployment 6A (aligns with D4 production): December 2015
  - Deployment 6B (aligns with D5 Production): December 2016

Gartner eManifest Project Assessment Executive Summary

Scope of the Assessment	Gartner has been asked to assess the eManifest recovery plan. This limits the analysis and possible recommendations.
Architecture Viability	The Architecture supports the realization of business benefits for eManifest. Non-functional requirements and the ability to meet them have yet to be confirmed.
Management of Risk	Past project delivery issues have been identified and risk mitigation actions developed. Executing those actions will be challenging.
Compressed delivery timelines	Project success will depend on prioritization of eManifest within the Agency. The magnitude of risk and change management is not fully appreciated.
Vendor Management	Procuring and managing complex outcome-based supply arrangements would prove very risky, given CBSA's current level of vendor management maturity. Executing new sourcing approaches may not yield the value being sought within eManifest's timelines.

Communications

- External Stakeholders
  - External stakeholders will be advised of the new delivery plan with dates
  - A full communications strategy will be developed once the plan is approved
- Internal Stakeholders
  - Present re-baselined plan to CIOB
  - Briefing to TB
- Minister's Office
  - Provide an update on eManifest status and plan, impacts on BtB commitments, external stakeholders and regulatory package
- PCO
  - Continue to update BtB team of eManifest status and impacts on BtB commitments
- Staff and Unions
  - Communication of the HR Strategy with staff and unions will begin once plan has been approved
- SSC
  - They are on the recovery team but will also be formally briefed on the approved plan

HR Strategy

- Project Development Lifecycle will see shifts in resource allocation and skillsets as the Project moves through Development and into Testing & Implementation

- Business and Systems Analysts will begin transitioning to Maintenance & to new project work through this fiscal and into 2015/2016

- Considering various external delivery assistance options for Business Intelligence (D6) along with the work done to date within the Agency.

Canada Border Services Agency | Agence des services frontaliers du Canada

eManifest Deployment Strategy Overview

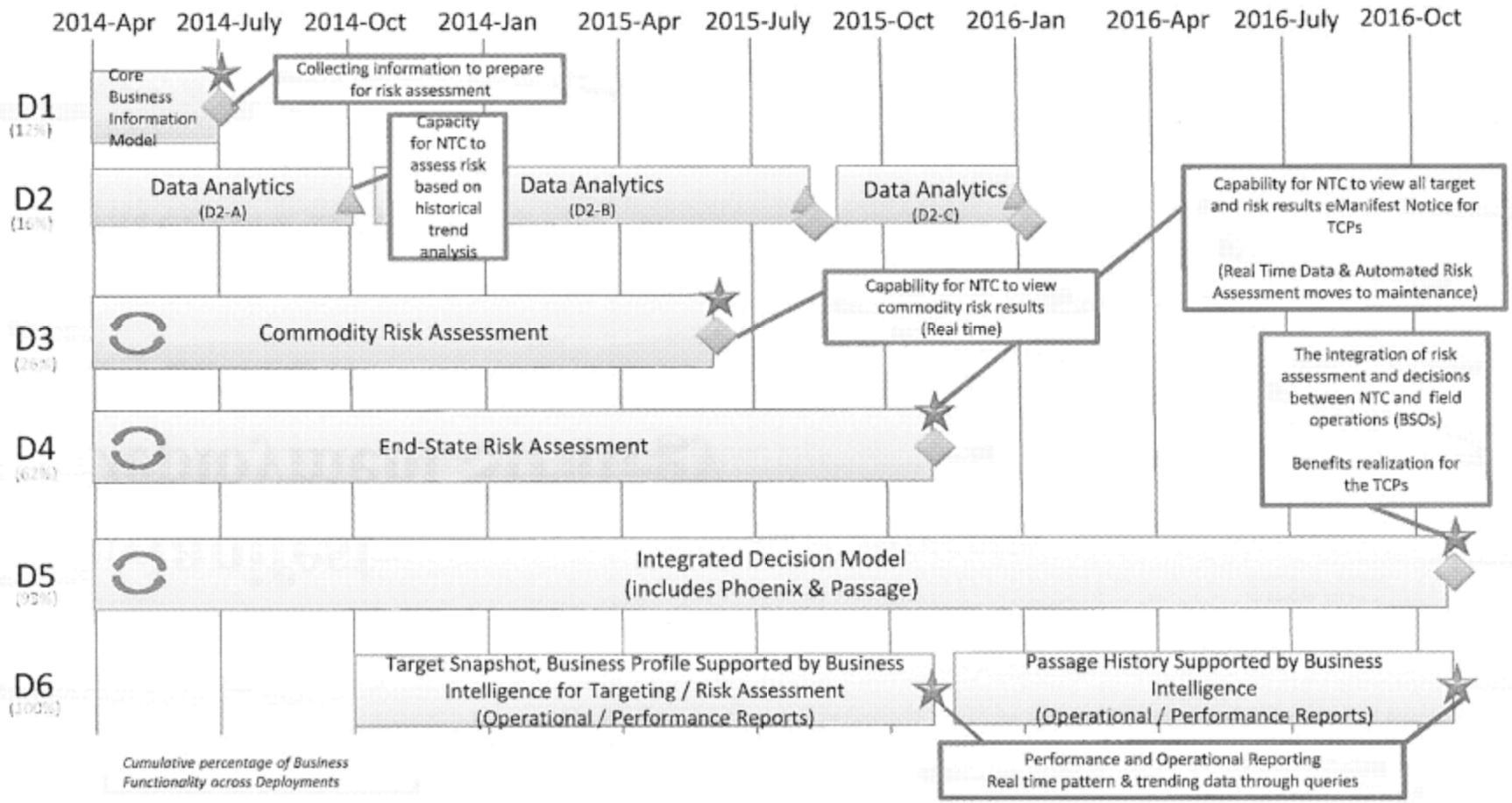
Executive Briefing

May 9, 2014
Version 1.0

PROTECTION • SERVICE • INTEGRITY

Canada

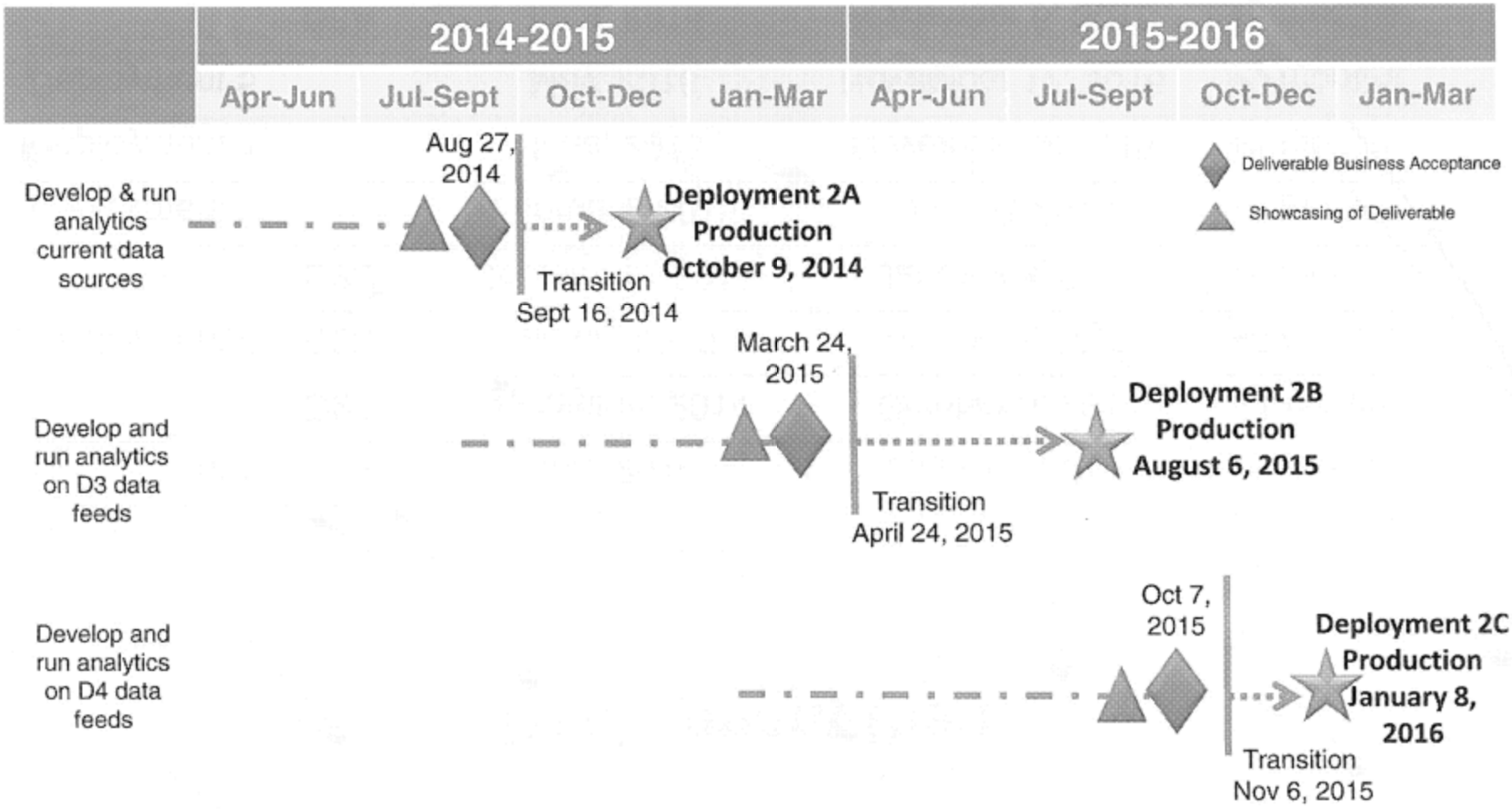
eManifest Proposed Deployment Strategy



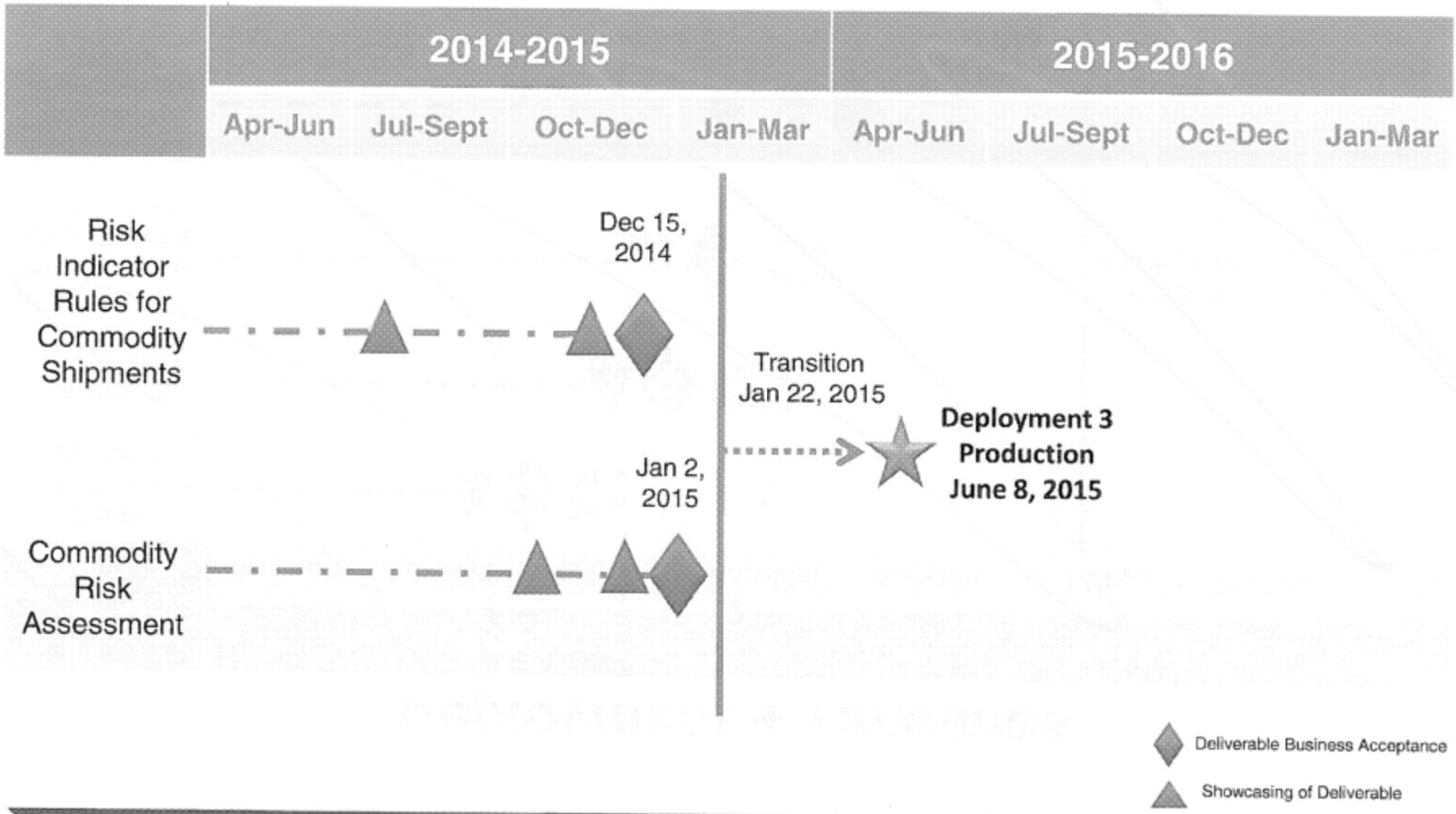
Deployment Plan

Deployment		Draft Date Identified in President Briefing	Rebaseline Date	Variance
Deployment 1		July, 2014	June 14, 2014	-1 month
Deployment 2	D2A	November, 2014	October 9, 2014	-1 month
	D2B	March, 2015	August 6, 2015	+5 months
	D2C	September, 2015	January 8, 2016	+4 months
Deployment 3		February, 2015	June 8, 2015	+4 months
Deployment 4		June, 2015	November 6, 2015	+5 months
Deployment 5		May, 2016	November 17, 2016	+6 months
Deployment 6	D6A	June, 2015	November 6, 2015	+5 months
	D6B	May, 2016	November 17, 2016	+6 months

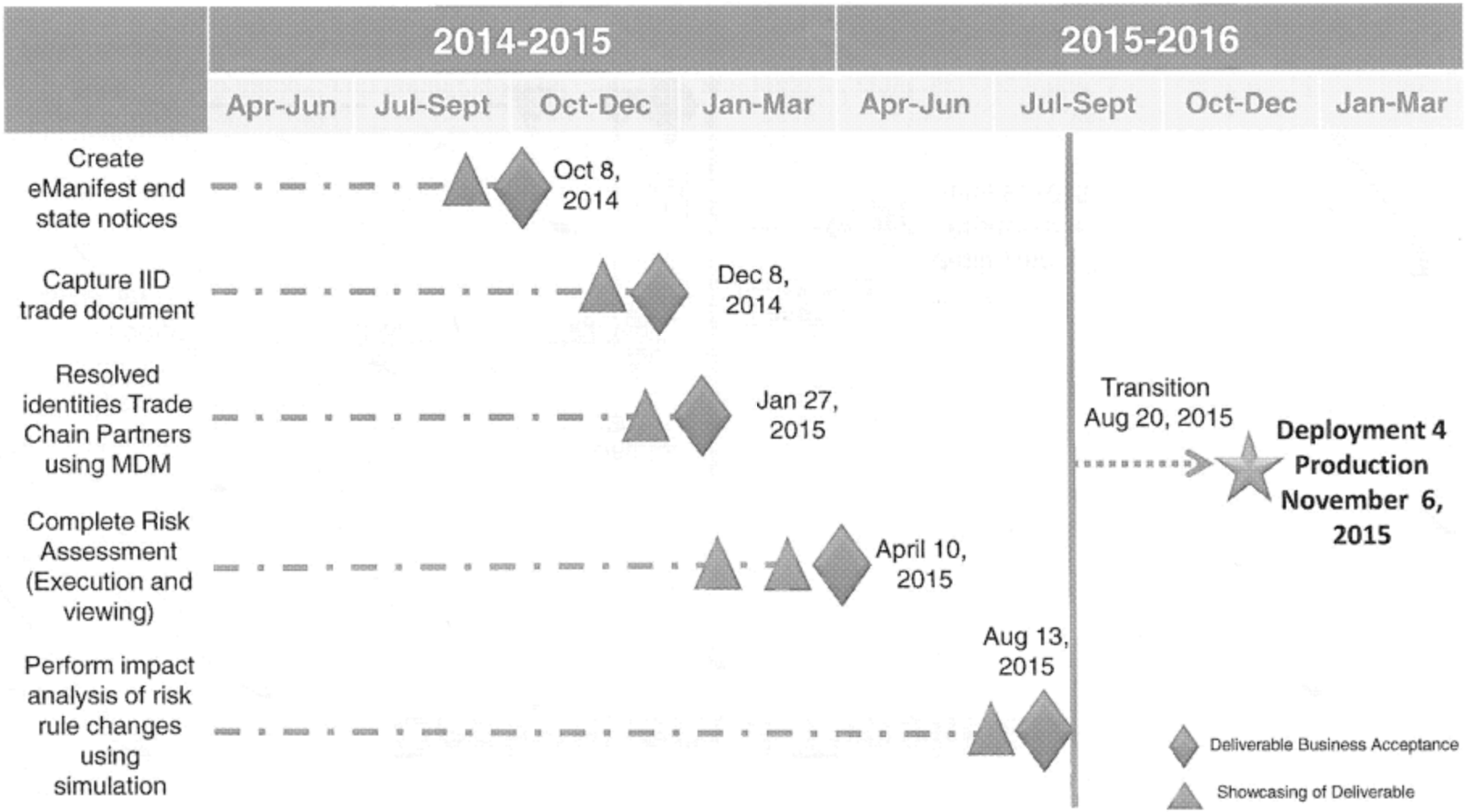
Deployment 2 Timelines



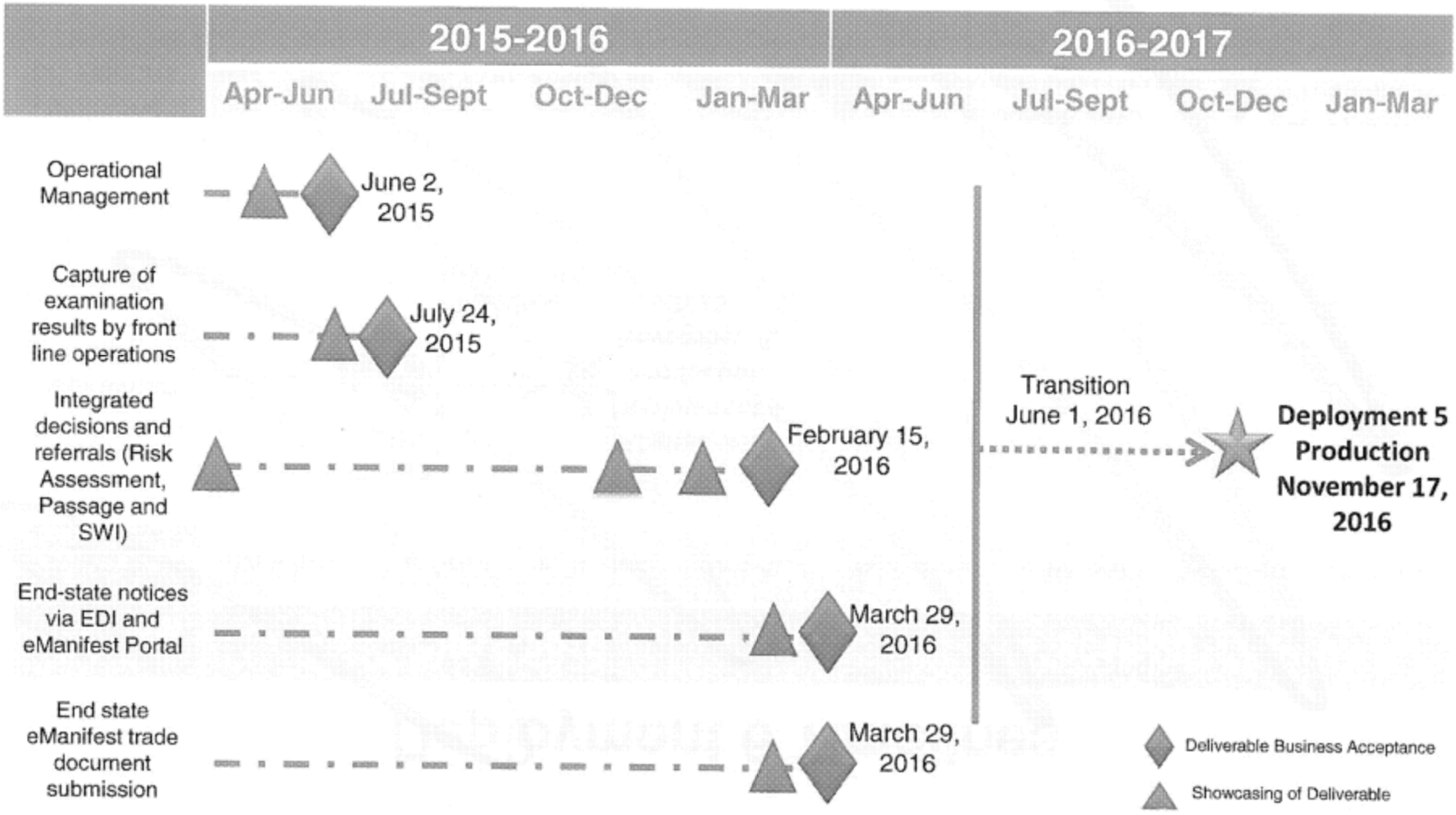
Deployment 3 Timelines



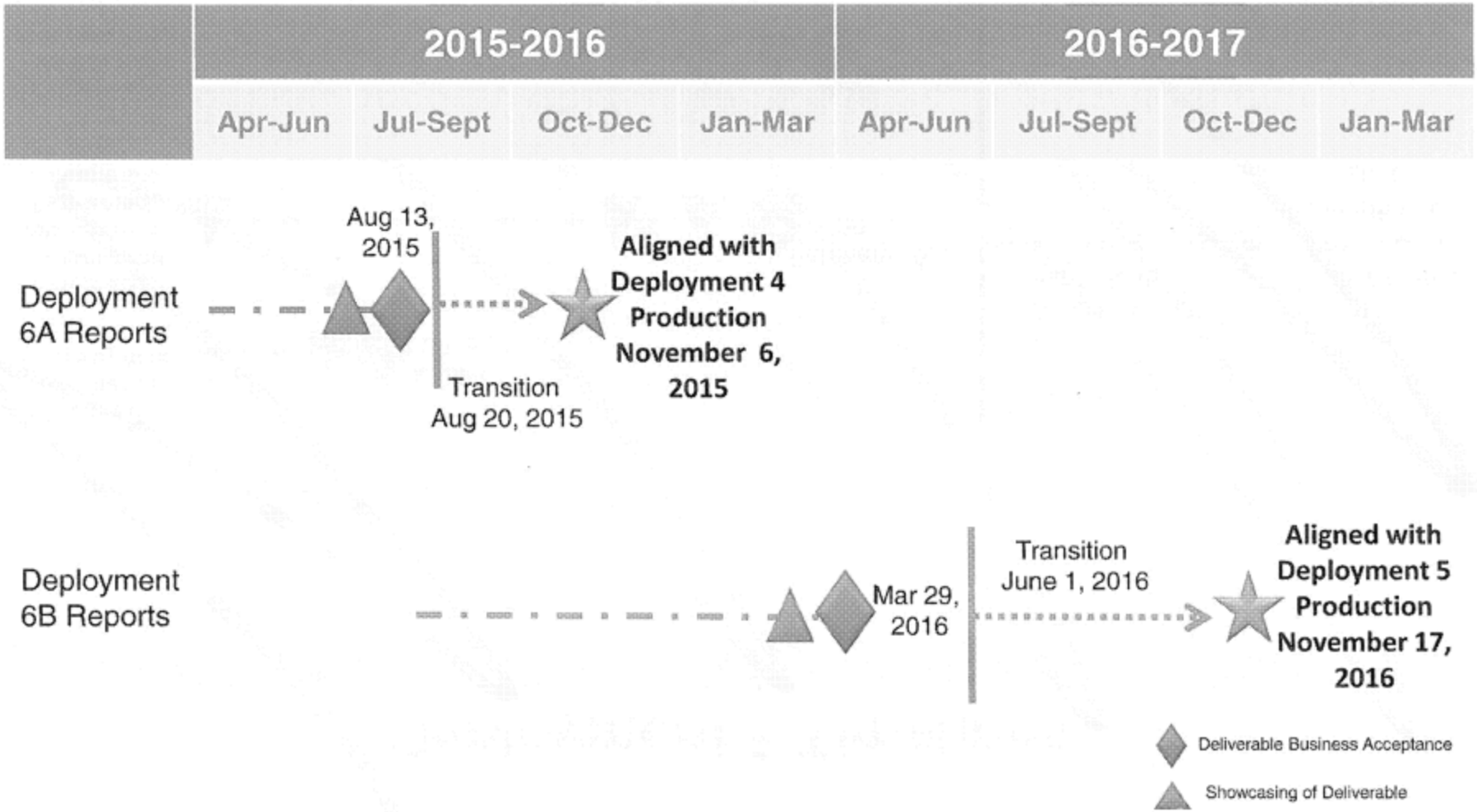
Deployment 4 Timelines



Deployment 5 Timelines



Deployment 6 Timelines



Back-Up Slides

Deployment 2 Deliverables

Deliverable	Showcase
Develop and execute analytical models based on Program / Operational priorities using current data sources.	• Run analytical models using Enterprise Data Warehouse appliance current data sources
Demonstrate that analytical models can support Targeting and Intelligence gathering using new data feeds.	• Run analytical models against current and new data sources from Deployment 3
Develop candidate risk indicators or modify existing indicators based on analysis / outcomes and new data feeds.	• Run analytical models against current and new data sources from Deployment 4

Deployment 2 Timelines

	Phase	Milestone	Date
Deployment 2A	Pre Transition	Develop & execute analytical models based on Program / Operational priorities using current data sources – Business Acceptance	August 27, 2014
	Transition	Deployment Transition Readiness	September 9, 2014
	Production	Deployment Production	October 9, 2014
Deployment 2B	Pre Transition	Demonstrate that analytical models can support Targeting and Intelligence gathering using new data feeds – Business Acceptance	March 24, 2015
	Transition	Deployment Transition Readiness	April 24, 2015
	Production	Deployment Production	August 6, 2015
Deployment 2C	Pre Transition	Develop candidate risk indicators or modify existing indicators based on analysis / outcomes and new data feeds – Business Acceptance	October 7, 2015
	Transition	Deployment Transition Readiness	November 6, 2015
	Production	Deployment Production	January 8, 2015

Deployment 3 Deliverables

Deliverable	Showcasing of Deliverable
Risk Indicator Rules for Commodity Shipments (Viewing, Authoring and Deploying using ODM)	• Cargo Commodity Risk Assessment • Shipment Commodity Risk Assessment
Commodity Risk Assessment	• Find the Commodity Risk • Shipment Versions

Deployment 3 Timelines

Phase	Milestone	Date
Pre Transition	Risk Indicator Rules for Commodity Shipments (Viewing, Authoring and Deploying using ODM) – Business Acceptance	December 15, 2014
	Commodity Risk Assessment – Business Acceptance	January 2, 2015
Transition	Deployment Transition Readiness	January 22, 2015
Production	Deployment Production	June 8, 2015

Deployment 4 Deliverables

Deliverable	Showcasing of Deliverable
Create eManifest end state notices (create and view)	• End State Notices
Capture IID trade document (create and view)	• Capture IID Trade Document
Resolved identities Trade Chain Partners using MDM	• Resolution of TCP Identity
Complete Risk Assessment (Execution and viewing)	• Comprehensive Risk Rule Set • Target Maintenance
Perform impact analysis of risk rule changes using simulation	• Impact of Rule Change (Simulation)

Deployment 4 Timelines

Phase	Milestone	Date
Pre Transition	Create eManifest end state notices (create and view) – Business Acceptance	October 8, 2014
	Capture IID trade document (create and view) – Business Acceptance	December 8, 2014
	Resolved identities Trade Chain Partners using MDM – Business Acceptance	January 27, 2015
	Complete Risk Assessment (Execution and viewing) – Business Acceptance	April 10, 2015
	Perform impact analysis of risk rule changes using simulation – Business Acceptance	August 13, 2015
Transition	Deployment Transition	August 20, 2015
Production	Deployment Production	November 6, 2015

Deployment 5 Deliverables

Deliverable	Showcasing of Deliverable
Operational Management	• Risk and Passage Superintendent
Capture of examination results by front line operations	• Integrated Front Counter and Secondary Decision
Integrated decisions and referrals (Risk Assessment, Passage and SWI)	• Integrated PIL Decision • Single-Window Integration • Transition from ACROSS/TITAN to eManifest
End-state notices via EDI and eManifest Portal	• External Client Communication
End state eManifest trade document submission	• External Client Communication

Deployment 5 Timelines

Phase	Milestone	Date
Pre Transition	Operational Management – Business Acceptance	June 2, 2015
	Capture of examination results by front line operations – Business Acceptance	July 24, 2015
	Integrated decisions and referrals (Risk Assessment, Passage and SWI) – Business Acceptance	February 15, 2016
	End-state notices via EDI and eManifest Portal – Business Acceptance	March 29, 2016
	End state eManifest trade document submission – Business Acceptance	March 29, 2016
Transition	Deployment Transition Readiness	June 1, 2016
Production	Deployment Production	November 17, 2016

Deployment 6 Deliverables

Deployment	Deliverable	Showcasing of Deliverable
Deployment 6A Reports *(aligns with D4 Production Release)*	Risk Assessment Reports	• Integrated Operational Business Intelligence Target Snapshot
	Operational Business Intelligence Reports to support Phoenix and Target Maintenance	
Deployment 6B Reports *(aligns with D5 Production Release)*	Passage Reports	• Passage Summary Report
	Operational Business Intelligence to support Passage	

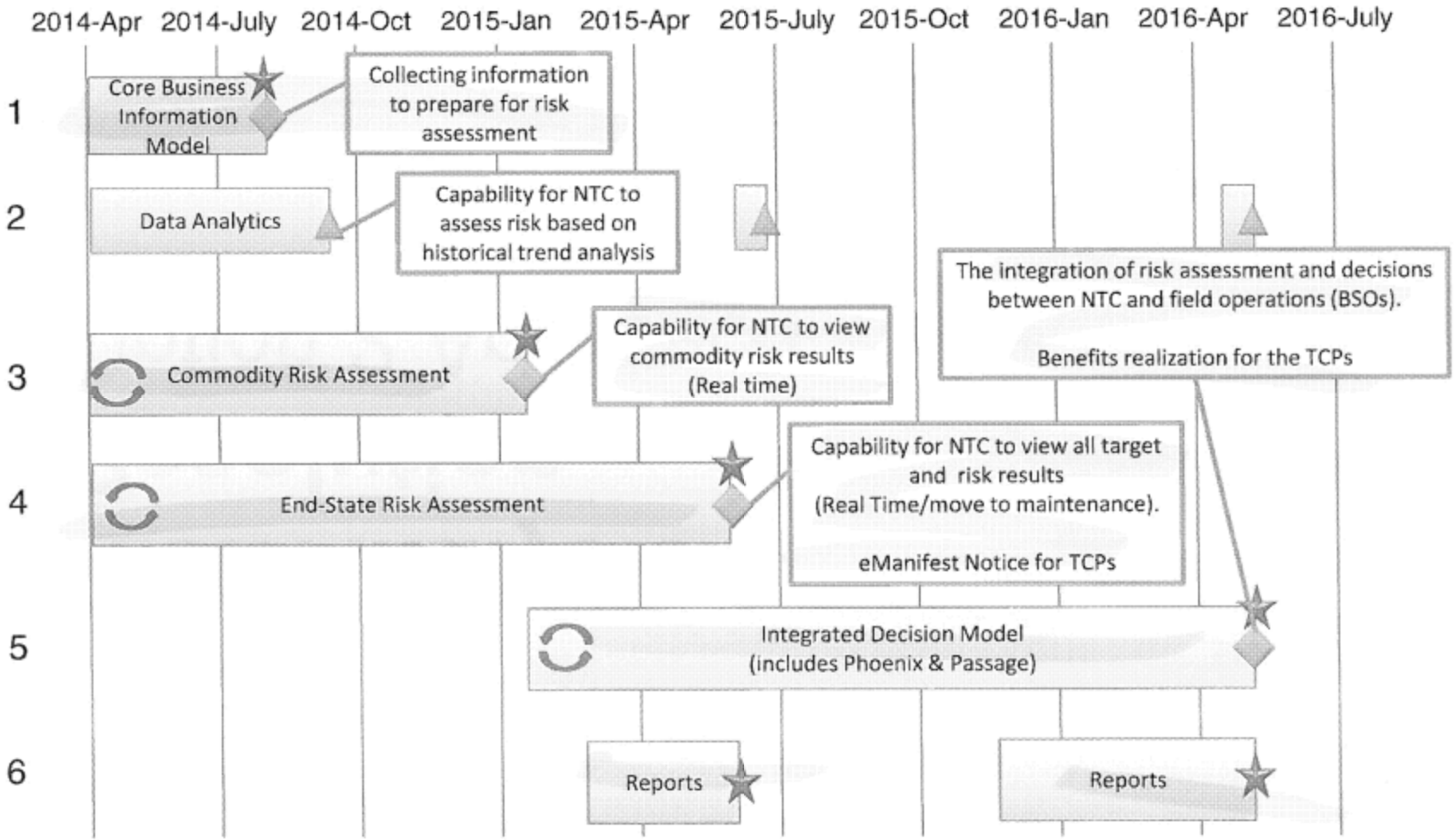
Deployment 6 Timelines

Phase	Milestone	Date
Pre Transition	Risk Assessment Reports – Business Acceptance	August 13, 2015
	Operational Business Intelligence Reports to support Phoenix and Target Maintenance – Business Acceptance	August 13, 2015
Transition	Deployment Transition Readiness	August 20, 2015
Production	Deployment Production *(aligns with D4 Production)*	November 6, 2015
Pre Transition	Passage Reports – Business Acceptance	March 29, 2016
	Operational Business Intelligence to support Passage – Business Acceptance	March 29, 2016
Transition	Deployment Transition Readiness	June 1, 2016
Production	Deployment Production *(aligns with D5 Production)*	November 17, 2016

Canada Border Services Agency | Agence des services frontaliers du Canada

Risk Analysis & Resolution

Option 2 Approach


2014-Apr
2014-July
2014-Oct
2015-Jan
2015-Apr
2015-July
2015-Oct
2016-Jan
2016-Apr
2016-July
1
2
3
4
5
6
Core Business Information Model
Collecting information to prepare for risk assessment
Data Analytics
Capability for NTC to assess risk based on historical trend analysis
The integration of risk assessment and decisions between NTC and field operations (BSOs).
Benefits realization for the TCPs
Commodity Risk Assessment
Capability for NTC to view commodity risk results (Real time)
Capability for NTC to view all target and risk results (Real Time/move to maintenance).
eManifest Notice for TCPs
End-State Risk Assessment
Integrated Decision Model (includes Phoenix & Passage)
Reports
Reports
Non STAR Release
STAR Release
Business Information Model Progression

Release Details

Core Business Information Model

(Stream #1)

Scope

- Transition existing trade document to establish the core business information model.
- Laying the foundation to prepare the trade documents to support risk assessment.
  - Commodity (Language Ware)
  - Address (QS-AVI)

Data Analytics

(Stream #2)

Scope

- Data warehouse (using existing Data Marts) operational on the new data warehouse appliance.
- Increased refresh frequency on a sub-set of the ACROSS data sources to support highway and rail movements.
- Copy the Business Information Model data (to be aligned with the progression dictated by the other eManifest streams).
- eManifest will provide a data exploitation expert within the NTC to assist existing intelligence officers.
- Any resulting actions or decisions will be performed in the existing applications (ACROSS/TITAN)

Commodity Risk Assessment

(Stream #3)

Scope

- Commodity Risk Assessment using the Business Information Model.
- Provide the risk assessment results to the Intelligence Officers in the NTC.
- Utilize the Business Analytics to identify new Commodity Risk Rules
- Provide the ability to modify the Commodity Risk rules in production outside of a normal release cycle.
  - ODM Decision Center
  - ODM Rules Execution Centre (RES)
- Provide a read-only view of the Business Information Model and Commodity Risk results (Phoenix).
- Transition of Business Information Model and commodity risk results to the data analytics environment.
- Record user audit information.

End-State Risk Assessment

(Stream #4)

Scope

- End-State Risk Assessment using the Business Information Model.
- Provide the risk assessment results to the Intelligence Officers in the NTC.
- Provide the ability to modify all Risk rules in production outside of a normal release cycle.
- Utilize the Business Analytics to identify new Risk Rules
- Provide the ability to utilize identity resolution to support Party based risk assessment:
  - MDM
- Provide the ability to manage and execute targets.
- Extending the Business Information Model to support all current trade document including IID.
- Provide an enhanced read-only view of the Business Information Model and Risk results (Phoenix) including user workload filters through the identification of low risk (ELR).
- Provide the ability to determine the operational impact of risk rule changes through simulation.
- Early delivery of eManifest end-state notices using legacy technology (ACROSS)
- Transition of Business Information Model and risk results to the data analytics environment.

Integrated Decision Model

(Stream #5)

Scope

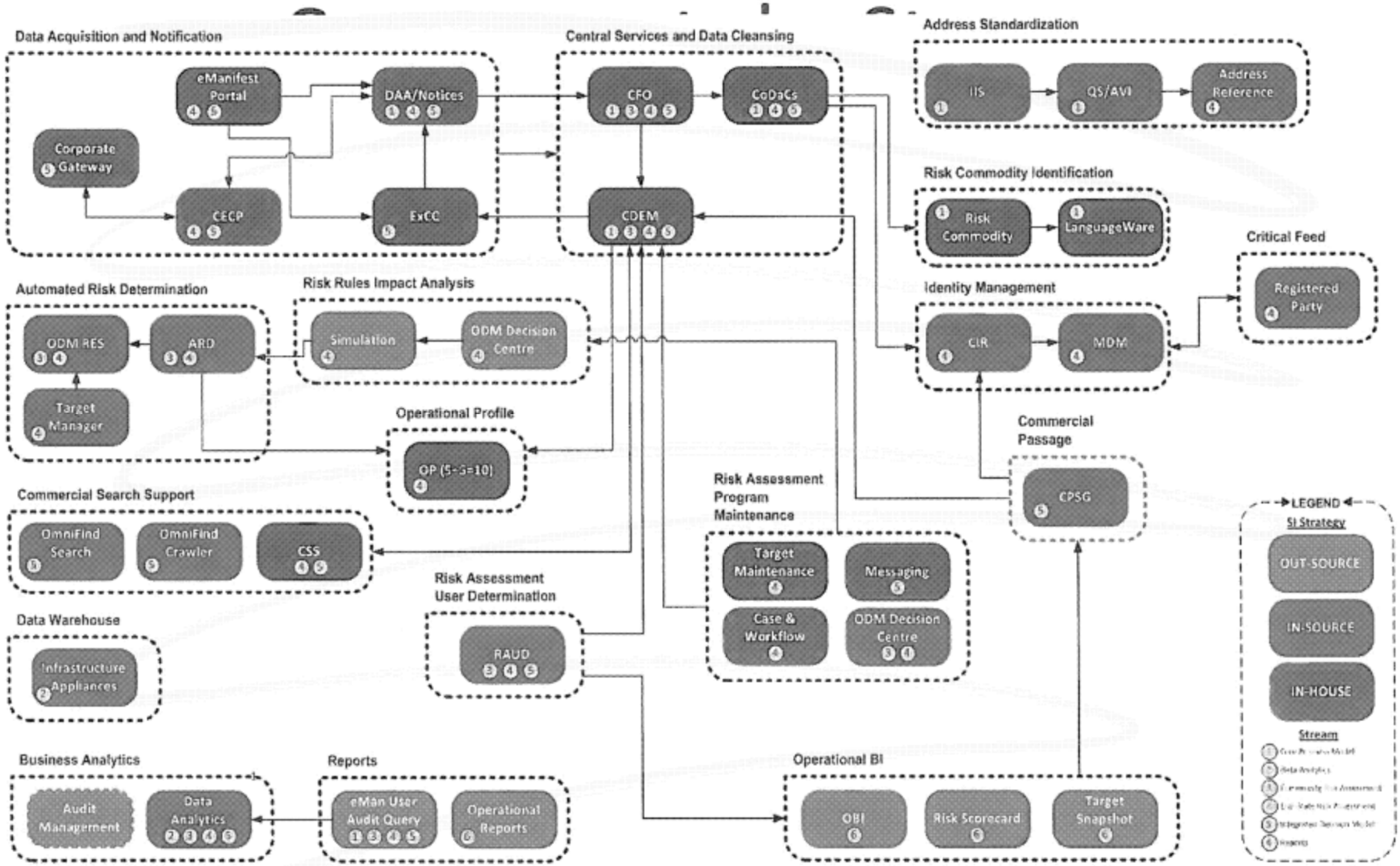
- Integrated Decisions using the end-state Business Information Model.
- Provide an integrated view of the Business Information Model and Risk Results to the front-line BSOs (Passage), the NTC targeting officers (Phoenix) and the Risk Assessment Program officers (RAPM).
- Provide the ability for the front-line BSOs (Passage) the NTC targeting officers (Phoenix) to make integrated operational decisions using the end-state Business Information Model.
- Extending the Business Information Model to support the end-state trade documents.
- Provide the ability to record referral details and examination results.
- Deliver eManifest notices using the end-state Business Information Model while still supporting the legacy notices.
- eManifest Portal enhanced to use the end-state Business Information Model.
- Legacy application (ACROSS/TITAN) downsized.
- Transition of Business Information Model, risk results and decisions to the data analytics environment.

Reports

(Stream #6)

Scope

- Risk Assessment Operational Reports
- Risk Assessment Management Reports
- Passage Operational Reports
- Passage Management Reports


Data Acquisition and Notification
eManifest Portal
DAA/Notices
Corporate Gateway
CECP
ExCC
Central Services and Data Cleansing
CFO
CoDaCs
CDEM
Address Standardization
IIS
QS/AVI
Address Reference
Risk Commodity Identification
Risk Commodity
LanguageWare
Critical Feed
Registered Party
Identity Management
CIR
MDM
Automated Risk Determination
ODM RES
ARD
Target Manager
Risk Rules Impact Analysis
Simulation
ODM Decision Centre
Operational Profile
OP (5-5=10)
Commercial Passage
CPSG
Commercial Search Support
OmniFind Search
OmniFind Crawler
CSS
Risk Assessment Program Maintenance
Target Maintenance
Messaging
Case & Workflow
ODM Decision Centre
Risk Assessment User Determination
RAUD
Data Warehouse
Infrastructure Appliances
LEGEND
SI Strategy
OUT-SOURCE
IN-SOURCE
IN-HOUSE
Stream
Business Analytics
Audit Management
Data Analytics
Reports
eMan User Audit Query
Operational Reports
Operational BI
OBI
Risk Scorecard
Target Snapshot